

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2003, 19:40:47 ; Search time 14833.9 Seconds
(without alignments)
17396.489 Million cell updates/sec

Title: US-09-724-126A-1

Perfect score: 6308

Sequence: 1 gccagaattcgccagcagg.....aattttgtattggtgtttt 6308

Scoring table:

IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:**

1: gb_ba:**

2: gb_hgt:**

3: gb_in:**

4: gb_om:**

5: gb_ov:**

6: gb_pat:**

7: gb_ph:**

8: gb_pl:**

9: gb_pr:**

10: gb_ro:**

11: gb_sts:**

12: gb_sy:**

13: gb_un:**

14: gb_vl:**

15: em_ba:**

16: em_fun:**

17: em_hum:**

18: em_in:**

19: em_mu:**

20: em_om:**

21: em_or:**

22: em_ov:**

23: em_pat:**

24: em_ph:**

25: em_pl:**

26: em_ro:**

27: em_sts:**

28: em_un:**

29: em_vl:**

30: em_hgt_hum:**

31: em_hgt_inv:**

32: em_hgt_other:**

33: em_hgt_mus:**

34: em_hgt_pin:**

35: em_hgt_rod:**

36: em_hgt_mam:**

37: em_hgt_vrt:**

38: em_sy:**

39: em_hgtgo_hum:**

40: em_hgtgo_mus:**

41: em_hgtgo_other:**

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	5250	83.2	5250	9	AY061886	AY061886 Homo sapi
2	5085.6	80.6	5128	9	AF525401	AF525401 Homo sapi
3	4435.4	70.3	6395	6	AR030784	AR030784 Sequence
4	4435.4	70.3	6395	6	AR121463	AR121463 Sequence
5	4435.4	70.3	6395	10	AF061555	AF061555 Mus muscu
6	3055.8	48.4	3059	6	AX714232	AX714232 Sequence
7	3055.8	48.4	3059	6	AK056441	AK056441 Homo sapi
8	2546.8	40.4	2550	6	BD156870	BD156870 Primer fo
9	2546.8	40.4	2550	9	AK027803	AK027803 Homo sapi
10	1378.6	21.9	5268	9	AY061884	AY061884 Homo sapi
11	1233.4	19.6	5265	10	AY061885	AY061885 Mus muscu
12	996.2	15.8	1001	6	AR030785	AR030785 Sequence
13	996.2	15.8	1001	6	AR121464	AR121464 Sequence
14	994.2	15.8	999	9	AF061556	AF061556 Homo sapi
15	981	15.6	6158	9	AB002347	AB002347 Human mRN
16	772.8	12.3	3502	10	BC031403	BC031403 Mus muscu
17	784.2	12.1	818	9	HX5521	Z47040 Human parti
18	651.6	10.3	2958	9	AK026998	AK026998 Homo sapi
19	647	10.3	712	6	BD149613	BD149613 Primer fo
20	515.2	8.2	2475	10	BC026391	BC026391 Mus muscu
21	503.8	8.0	16518	9	AC090514	AC090514 Homo sapi
22	487.8	7.7	2512	9	AK026948	AK026948 Homo sapi
23	354	5.6	190727	9	AC068724	AC068724 Homo sapi
24	314.8	5.0	1368	10	BC025617	BC025617 Mus muscu
25	290.4	4.6	235978	2	AC094209	AC094209 Rattus no
26	287.6	4.6	181602	10	AL935168	AL935168 Mouse DNA
27	279.4	4.4	61896	2	AC090992	AC090992 Homo sapi
28	246.8	3.9	807	6	BD079625	BD079625 Cancer-as
29	224	3.6	5027	10	AK122254	AK122254 Mus muscu
30	216.8	3.4	6381	3	AY094815	AY094815 Drosophil
31	215.2	3.4	910	10	MM1UBR2	AF067372 Mus muscu
32	215.2	3.4	168200	2	AC120631	AC120631 Rattus no
33	215.2	3.4	189946	10	AL844548	AL844548 Mouse DNA
34	212.2	3.4	107304	2	AC016274	AC016274 Homo sapi
35	212.2	3.4	164468	2	AC021899	AC021899 Homo sapi
36	191.6	3.0	800	6	BD079626	BD079626 Cancer-as
37	180.8	2.9	164468	2	AC021899	AC021899 Homo sapi
38	160.6	2.5	3980	9	HS1UBR4	AF067383 Homo sapi
39	160.2	2.5	3411	9	AK090874	AK090874 Homo sapi
40	159	2.5	455	6	BD071823	BD071823 Secreted
41	157.6	2.5	594	11	G79301	G79301 S210P6045RG
42	150.2	2.4	107304	2	AC016274	AC016274 Homo sapi
43	132	2.1	28834	2	AC017982	AC017982 Drosophil
44	132	2.1	152545	3	AC010922	AC010922 Drosophil
45	132	2.1	179016	3	AC018489	AC018489 Drosophil

ALIGNMENTS

RESULT 1
AY061886
LOCUS
DEFINITION Homo sapiens ubiquitin ligase E3 alpha-I mRNA, complete cds.
ACCESSION AY061886
VERSION AY061886.1 GI:27434483
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 5250)
AUTHORS Han,H.Q. and Kwak,K.
TITLE Full-length human ubiquitin ligase E3 alpha-I (E3 alpha)
JOURNAL Unpublished

REFERENCE	2 (bases 1 to 5250)
AUTHORS	Han,H.Q. and Kwak,K.
TITLE	Direct Submission
JOURNAL	Submitted (05-NOV-2001) Research, Amgen Inc., One Amgen Center Drive, Thousand Oaks, CA 91320, USA
FEATURES	Location/Qualifiers
	1..5250
source	/organism="Homo sapiens"
	/mol_type="mRNA"
CDS	/db_xref="taxon:9606"
	1..5250
	/codon_start=1
	/product="ubiquitin ligase E3 alpha-I"
	/protein_id="AAL32103.1"
	/db_xref="gi:2743448"
	/translation="MADEEAGQGERMEISAEELPQTPQRLASWMDQOVDFVTFATLHLHA QLVPEIFYAEMPDLPEKQESVQMSFTPLEWYLFQEDPDICLEKLKHSAGATQCLGRV FKSGETTYSCRCADIPTCVLMDCFQDSVHKHNRHKMTSTGGGFCDCGDEAWKGT PFCVNEHPGAGTITKENSRLPNEEVYQVARKIFPSVIKYIVEMTIEWEEKELPPLEQ IRKNEYCYVLFDNDEHSHYDVHYSIQALDCELAELAQHTTAIDKERRAVKAGAY AACAEKEDIKSHSENSQRPHEVLYHSEIMAHQKFLALRGLSMWNKIMSYSDSFQI FQACLRPEPDSNPCLISRLMLDAKLYGARKILHELIFSFFMEMEYKKLFAMEF VKYIKOLQEYISDDHRSITATLSVQMTVPTLARHLIEQNVTVITETLLEVL EYLDRKNFQYSDSKLGRVAVTCDLKYILISKPTLWTEPLRMQFLEGPESFLKI LTCQMGEEIRIQGVGHIIEVDPDWEAAIALQMLKNILMFQEWACADELLLVATKE CHAVMRCSTFSSSTKTVQSCGHSTKSYKVSSEDLVSIHLPLSTLAGHVRUSR LGAVALHFEFSEFQVRLVVEYPLRCLVLAQVVAEMRRNGLSLISQVIFYQDVK CREEMDKDIIIMQLIGASLMDPNKFLLLVQRYELAEAFNKTIISTKDQDLIKOYNTLI EEMLOLVIIYIGERYVPGVGNVTKEBVTMREITHLICIRPMHSAIAKLNPENNET GLENLVINKVATPKPCGVSGVYELKDESLKOFNMYFYHYSKTQHSKAEHMOKREKO ENKDEALPPPPPEPCPAPSKVINLNCDIMYILKTVFERAIDTSSNLWTEGMLQMA FHILAALLEGKQOLQKADPEEVTDFYHKASGLSSAMNIMLLEKJLQIPOLEGOK DITLWLOMFTVYKRLKESCLVATTSGSESKNDEITHDKEAKERKKAEEARLHR QKTMQMSALQKNFIETHKLMYDNTSEMPCKEDSIMEEESTPAVSYSRIALGPKRGP SVPEKEVLTKICQEQEVQKIENNAVLGACVOKSTALQHGKLTSELSEALDPLFM DPLAYLVTTCGSGGVHVAQWQKYFEAVQLSSQORIHVDLPDESSEYLCPLCKSLC NTVPIILPQPKINSENADALQILTLARITQTVLARISGVNIRAKGENPIPIFFN QMGSDSTLFHSLISGVSESSIKNSIKEMVILFATITIRIGLKYPPDPEDRPVPEL TQWCAFTIQAILENLGDGKPLFGALQNRHNGKALMQFAVAQRITCPQVILKHL VRLLSVLPNIXKESDTPCLLSIDFLHVGAVLAFPSLYWDDPDVLQPSVSSYNNHL YLFLLTITMAHMLQILITVDTGTLPAQWQEDSEASASSFFFAEISQYSGSIGCDPLG YLWVSLDNGITPTPLKCAALFHYLLGVTPPEELTNSAEGYSALCSYLSLCLNLAFL LFOEYVKTNPILQWCAQDALLNCLKQNTVYVYPRKNSLLELPDDYSLCLNLAASH FPCPRANDDERKHPVLCFGGAILCSQNICQBEIVNGEEVGCACIFIALHAGGAGVCFIL KIKERARVIEVQKARGCAVPAPVLDYEIGETDPLGLKRNPLHLSRERYRKLHLWQWQHC IIEEIRARSETQNMQLPFGFNQQLI"
BASE COUNT	1576 a 1026 c 1180 g 1468 t
ORIGIN	
Query Match	83.2%; Score 5250; DB 9; Length 5250;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 5250; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Qy	696 ATGCGGACGAGGAGGCTGGAGGTACTTGAGAGGATGGAAATCAGCGCGGAGTTTACCCCG 755
Db	1 ATGCGGACGAGGAGGCTGGAGGTACTTGAGAGGATGGAAATCAGCGCGGAGTTTACCCCG 60
Qy	756 ACCCCTCAGCGTCTGGCATCTTGGTGGATCAGCAAGTTGATTTTATACTGCTTTCTTG 815
Db	61 ACCCCTCAGCGTCTGGCATCTTGGTGGATCAGCAAGTTGATTTTATACTGCTTTCTTG 120
Qy	816 CATCATTTGGCAAAATTTGGTGCCAGAAATTTACTTTGCTGAATGGACCCAGACTTGGAA 875
Db	121 CATCATTTGGCAAAATTTGGTGCCAGAAATTTACTTTGCTGAATGGACCCAGACTTGGAA 180
Qy	876 AAGCAGGAGGAAAGTGTFACAAATGTCAATATTCACTCCACTGGAATGGTACTTATTGGGA 935
Db	181 AAGCAGGAGGAAAGTGTFACAAATGTCAATATTCACTCCACTGGAATGGTACTTATTGGGA 240
Qy	936 GAGATCCAGATATTTGCTTTAGAGAAATTAAGCACAGTGGAGCATTTTACGCTTTGTGGG 995
Db	241 GAAGATCCAGATATTTGCTTTAGAGAAATTAAGCACAGTGGAGCATTTTACGCTTTGTGGG 300

Qy	996	AGGTTTTC	AAAAGTGGAGAGACAAACCTATTC	TTCGAGGATTCGCAATTCATGATCAACA	105
Db	301	AGGGTTTTC	AAAAGTGGAGAGACAAACCTATTC	TTCGAGGATTCGCAATTCATGATCAACA	360
Qy	1056	TGTTACTCTG	TATGGACTGCTCCAGGACAGTGTTCAT	AAAAATCATCGTTTACAAGATG	1115
Db	361	TGTTACTCTG	TATGGACTGCTCCAGGACAGTGTTCAT	AAAAATCATCGTTTACAAGATG	420
Qy	1116	CATACTTCT	ACTGGAGAGGGTTCGTGACTGTGGAGACACAGAGGCATGGA	AAACTGGC	1175
Db	421	CATACTTCT	ACTGGAGAGGGTTCGTGACTGTGGAGACACAGAGGCATGGA	AAACTGGC	480
Qy	1176	CCTTTTTGT	TAATCATGAACCTGGAGAGCAGGTACTATAAAGAGAAATTCACGCTGT		1235
Db	481	CCTTTTTGT	TAATCATGAACCTGGAGAGCAGGTACTATAAAGAGAAATTCACGCTGT		540
Qy	1236	CCGTTGAAT	GAAGAGTTAATGTCCAAGCCAGGAAATATTTCC	TTCAGTGATAAATAT	1295
Db	541	CCGTTGAAT	GAAGAGTTAATGTCCAAGCCAGGAAATATTTCC	TTCAGTGATAAATAT	600
Qy	1296	GTCTGTAGA	AAATGACTATATGGGAAGAGGAAAGAACTGCCTCTG	AACTCCAGATAAGG	1355
Db	601	GTCTGTAGA	AAATGACTATATGGGAAGAGGAAAGAACTGCCTCTG	AACTCCAGATAAGG	660
Qy	1356	GAGAAAAAT	GAAAGATACTATTTGTGCTCTTTTCAATGATGAACACCACTTCATATGACCA	C	1415
Db	661	GAGAAAAAT	GAAAGATACTATTTGTGCTCTTTTCAATGATGAACACCACTTCATATGACCA	C	720
Qy	1416	GTCAATATAC	AGCCTACAAAGAGCTCTTGACTGTGAGCTCGAGAGGCCAGTTGCATACC		1475
Db	721	GTCAATATAC	AGCCTACAAAGAGCTCTTGACTGTGAGCTCGAGAGGCCAGTTGCATACC		780
Qy	1476	ACTGCCATT	GACAAAGAGGTCGCGGCTGTTTAAAGCGGAGCTTATGCTGCTGGCAG		1535
Db	781	ACTGCCATT	GACAAAGAGGTCGCGGCTGTTTAAAGCGGAGCTTATGCTGCTGGCAG		840
Qy	1536	GAAGCAAGG	AAAGATATAAAGAGTCATTTCAGAAAATGCTCTCAACATCCACTTCATGTA		1595
Db	841	GAAGCAAGG	AAAGATATAAAGAGTCATTTCAGAAAATGCTCTCAACATCCACTTCATGTA		900
Qy	1596	GAAGTATTTA	CACTCAGAGATTATGGCTCATCAGAAAATTTGCTTTGCGCTCTTGTTCTCTGG		1655
Db	901	GAAGTATTTA	CACTCAGAGATTATGGCTCATCAGAAAATTTGCTTTGCGCTCTTGTTCTCTGG		960
Qy	1656	ATGAACAAA	ATTTAGCTATTTCAAGTGACTTTTAGCGAGATCTTTTCCCAAGCATGCTT		1715
Db	961	ATGAACAAA	ATTTAGCTATTTCAAGTGACTTTTAGCGAGATCTTTTCCCAAGCATGCTT		1020
Qy	1716	AGAGAAGAAC	CTGACTCGGAAATCCCTGTCTCATAGCAGGTTAATGCTTTGGGATGCA		1775
Db	1021	AGAGAAGAAC	CTGACTCGGAAATCCCTGTCTCATAGCAGGTTAATGCTTTGGGATGCA		1080
Qy	1776	AAGCTTTTAA	AGGTGCCGTAAAGATCGTTTCATGAAATTTGATCTTCACGAGTTTTTTTATG		1835
Db	1081	AAGCTTTTAA	AGGTGCCGTAAAGATCGTTTCATGAAATTTGATCTTCACGAGTTTTTTTATG		1140
Qy	1836	GAGATGGA	ATAACAAAACCTTTTGGCTATGGAATTTGTGAAGTATTATAAACAACTGCAG		1895
Db	1141	GAGATGGA	ATAACAAAACCTTTTGGCTATGGAATTTGTGAAGTATTATAAACAACTGCAG		1200
Qy	1896	AAAGAAATAT	ATATACGTGATGATCATGACAGAGTATCTCTATACTGCACCTTTTCAGTTTCAG		1955
Db	1201	AAAGAAATAT	ATATACGTGATGATCATGACAGAGTATCTCTATACTGCACCTTTTCAGTTTCAG		1260
Qy	1956	ATGTTTACTG	TTTCTCTCGGCTGCACATCTTATTTGAAGCAGAGATGTTATCTCTGTCTC		2015
Db	1261	ATGTTTACTG	TTTCTCTCGGCTGCACATCTTATTTGAAGCAGAGATGTTATCTCTGTCTC		1320
Qy	2016	ATTACTGAA	CTCTGCTAGAGTTTTTACCTGAGTACTTGGACAGGAAACAATAAATTCAC		2075
Db	1321	ATTACTGAA	CTCTGCTAGAGTTTTTACCTGAGTACTTGGACAGGAAACAATAAATTCAC		1380
Qy	2076	TTCCAGGGT	TATAGCCAGGACAAATTTGGGAAGAGTATATG	CAGTAAATATGTGACCTAAAG	2135

QY	4296	AAATCTCTGTGCAATACTGTGATCCCCATTTATTCCTTTGCAACCTCAAAAGATAAACAGT	4355
Db	3601	AAATCTCTGTGCAATACTGTGATCCCCATTTATTCCTTTGCAACCTCAAAAGATAAACAGT	3660
QY	4356	GAGAAATGCAGATGCTCTTGCTCAACTTTTGACCTTGGCAGGGTGGATACAGACTGTCTCG	4415
Db	3661	GAGAAATGCAGATGCTCTTGCTCAACTTTTGACCTTGGCAGGGTGGATACAGACTGTCTCG	3720
QY	4416	GCCAGAAATATCAGGTTATAATATAAGACATGCTAAAGGAGAAAACCAATTCCTATTTC	4475
Db	3721	GCCAGAAATATCAGGTTATAATATAAGACATGCTAAAGGAGAAAACCAATTCCTATTTC	3780
QY	4476	TTTAAATCAAGGAATGGGAGATTCTACTTTGGAGTTCCATTCATCTCGAGTTTGGCGTT	4535
Db	3781	TTTAAATCAAGGAATGGGAGATTCTACTTTGGAGTTCCATTCATCTCGAGTTTGGCGTT	3840
QY	4536	GAGTCTTCGAAATAAATATTCAAATAGCATCAAGGAATGGTTATCTCTTTGGCCACAACA	4595
Db	3841	GAGTCTTCGAAATAAATATTCAAATAGCATCAAGGAATGGTTATCTCTTTGGCCACAACA	3900
QY	4596	ATTATATAGAAATGGATTGAAAGTGCCACCTGATGAAAGGGATCCTCGAGTCCCATGCTG	4655
Db	3901	ATTATATAGAAATGGATTGAAAGTGCCACCTGATGAAAGGGATCCTCGAGTCCCATGCTG	3960
QY	4656	ACCTGGAGCACCTCCGCTTTCACATACCAGGCAATTTGAAATCTATTGGGAGATGAAGGA	4715
Db	3961	ACCTGGAGCACCTCCGCTTTCACATACCAGGCAATTTGAAATCTATTGGGAGATGAAGGA	4020
QY	4716	AAACCTCTGTTTGGAGCAGCTTCAAATAGCAGCATAAATGGTCTGAAAGCATTTAATGCAG	4775
Db	4021	AAACCTCTGTTTGGAGCAGCTTCAAATAGCAGCATAAATGGTCTGAAAGCATTTAATGCAG	4080
QY	4776	TTTGCAGTTGCACAGAGGATTACCTGTCCTCAGGTCTGTATACAGAAACATCTGGTTTCTG	4835
Db	4081	TTTGCAGTTGCACAGAGGATTACCTGTCCTCAGGTCTGTATACAGAAACATCTGGTTTCTG	4140
QY	4836	CTTCTATACAGTTGTTCTTCTAACATAAAATCAGAAGATACACCATGCCCTTCTGTCTATA	4895
Db	4141	CTTCTATACAGTTGTTCTTCTAACATAAAATCAGAAGATACACCATGCCCTTCTGTCTATA	4200
QY	4896	GATCTGTTTCATGTTTGGTGGGTGCTGTGTAGCATTTCCCATCTTGTTATTTGGGATGAC	4955
Db	4201	GATCTGTTTCATGTTTGGTGGGTGCTGTGTAGCATTTCCCATCTTGTTATTTGGGATGAC	4260
QY	4956	CCTGTTGATCTGCAGCCCTTCTCAGTTAGTTCCTTCCCTATAACCAACCTTTATCTCTCCAT	5015
Db	4261	CCTGTTGATCTGCAGCCCTTCTCAGTTAGTTCCTTCCCTATAACCAACCTTTATCTCTCCAT	4320
QY	5016	TTGATACCAATGGCACACATGCTTCAGATACTACTTACATAGACACAGGCCCTACCCCTT	5075
Db	4321	TTGATACCAATGGCACACATGCTTCAGATACTACTTACATAGACACAGGCCCTACCCCTT	4380
QY	5076	GCTCAGGTTCAAGAAGACAGTGAAGAGGCTCATTCGCCCATCTTCTTCTTTTGCAGAATTT	5135
Db	4381	GCTCAGGTTCAAGAAGACAGTGAAGAGGCTCATTCGCCCATCTTCTTCTTTTGCAGAATTT	4440
QY	5136	TCTCAATATACAGTGGCTCCATTTGGGTGTGATATTCCTGGCTGGTATTTTGTGGGTCTCA	5195
Db	4441	TCTCAATATACAGTGGCTCCATTTGGGTGTGATATTCCTGGCTGGTATTTTGTGGGTCTCA	4500
QY	5196	CTGAAGAATGGCATCACCCCTTATCTTCGCTGTGCTCATTTGTTTTTCCACTATTTCCTT	5255
Db	4501	CTGAAGAATGGCATCACCCCTTATCTTCGCTGTGCTCATTTGTTTTTCCACTATTTCCTT	4560
QY	5256	GGGCTAACTCCGCTTGAGGAACATGCATACCAATTTCTGCAGAAGGAGAGTACAGTGCACTC	5315
Db	4561	GGGCTAACTCCGCTTGAGGAACATGCATACCAATTTCTGCAGAAGGAGAGTACAGTGCACTC	4620
QY	5316	TGTAGCTATCTATCTTTACCTACAAATTTGTTTCTCTGCTCTTCTCCAGGAATATTTGGGATPACT	5375
Db	4621	TGTAGCTATCTATCTTTACCTACAAATTTGTTTCTCTGCTCTTCTCCAGGAATATTTGGGATPACT	4680

QY	5376	GTAGGCCCTTGCCTCCAGAGGTGGTGTGCAGATCCTGCCCTTACTAAACTGTTTGAAGCAA	5433
Db	4681	GTAGGCCCTTGCCTCCAGAGGTGGTGTGCAGATCCTGCCCTTACTAAACTGTTTGAAGCAA	4740
QY	5436	AAAACACCGTGGTCAGGTACCCCTAGAAAAGAAATAGTTTGATAGAGCTTCCTGTATGAC	5495
Db	4741	AAAACACCGTGGTCAGGTACCCCTAGAAAAGAAATAGTTTGATAGAGCTTCCTGTATGAC	4800
QY	5496	TATAGCTGCCCTCCCTGAATCAAGCTTCTCATTTCCAGGTGCCACCGCTCTGCAGATGATGAG	5555
Db	4801	TATAGCTGCCCTCCCTGAATCAAGCTTCTCATTTCCAGGTGCCACCGCTCTGCAGATGATGAG	4860
QY	5556	CGAAGCATCCTGTCCCTCTGCCTTTTCTGTGGGCTATACTATGTTCTCAGAACATTTGC	5615
Db	4861	CGAAGCATCCTGTCCCTCTGCCTTTTCTGTGGGCTATACTATGTTCTCAGAACATTTGC	4920
QY	5616	TGCAGAGAAATGTGAACGGGGAAGAGGTTGGAGCTTGCATTTTTCAGGCATTCACCTGT	5675
Db	4921	TGCCAGGAAATTTGTGAACGGGGAAGAGGTTGGAGCTTGCATTTTTCAGGCATTCACCTGT	4980
QY	5676	GGAGCCGGAGTCTGCATTTTCTTAAATAATCAGAAATGCCGAGTGGTCTCTGGTTGAAGGT	5735
Db	4981	GGAGCCGGAGTCTGCATTTTCTTAAATAATCAGAAATGCCGAGTGGTCTCTGGTTGAAGGT	5040
QY	5736	AAAGCCAGAGGCTGTGCCATCCAGCTCCTTTACTTGGATGAATATGAGAAACAGACCCCT	5795
Db	5041	AAAGCCAGAGGCTGTGCCATCCAGCTCCTTTACTTGGATGAATATGAGAAACAGACCCCT	5100
QY	5796	GGCCTGAAGAGGGGCAACCCCTTCATTTATCTCGTGAGCGGTATCGGAAGCTCCATTTG	5855
Db	5101	GGCCTGAAGAGGGGCAACCCCTTCATTTATCTCGTGAGCGGTATCGGAAGCTCCATTTG	5160
QY	5856	GTCTGGCAACCACTGCTATTATAGAAGAGATTCTCTAGGAGCCAAAGAGACTAATCAGATG	5915
Db	5161	GTCTGGCAACCACTGCTATTATAGAAGAGATTCTCTAGGAGCCAAAGAGACTAATCAGATG	5220
QY	5916	TTATTGGATTCAACTGGCAGTCTACTGTGA	5945
Db	5221	TTATTGGATTCAACTGGCAGTCTACTGTGA	5250
RESULT 2			
AF525401			
LOCUS	AF525401	5128 bp mRNA linear	PRI 01-JAN-2003
DEFINITION	Homo sapiens UB1 E3a ligase mRNA, partial cds.		
ACCESSION	AF525401		
VERSION	AF525401.1	GI:27451603	
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
	1 (bases 1 to 5128)		
	Dgany,O., Aviadan,N., Delaunay,J., Krasnov,T., Shalmon,L.,		
	Shalev,H., Edelitz-Markus,T., Kapelushnik,J., Cattani,D.,		
	Pariente,A., Tulliez,M., Cretien,A., Schischmanoff,P.O.,		
	Iolascon,A., Fibach,E., Koren,A., Rossler,J., Le Merrier,M.,		
	Yaniv,I., Zalizov,R., Ben-Asher,E., Glender,I., Lancet,D.,		
	Beckmann,J.S. and Tamary,H.		
	Congenital dyserythropoietic anemia type I is caused by mutations		
	in codanin-1		
TITLE	Am. J. Hum. Genet. 71 (6), 1467-1474 (2002)		
JOURNAL	22340442		
MEDLINE	12434312		
PUBMED	2 (bases 1 to 5128)		
REFERENCE	Ben-Asher,E.		
AUTHORS	Direct Submission		
TITLE	Submitted (27-JUN-2002) Molecular Genetics, The Weizmann Institute		
JOURNAL	of Science and the Schneider Medical Center, Hertzfel Street,		
	Rehovot 76100, Israel		
FEATURES	Location/Qualifiers		
source	1. .5128		
	/organism="Homo sapiens"		

QY	2247	ATCCGAAGACAGGTTGGCAACACATTTGAAGTGGATCGATCGGAGGCTGCCATTGCT	2306
DB	1501	ATCCGAAGACAGGTTGGCAACACATTTGAAGTGGATCGATCGGAGGCTGCCATTGCT	1560
QY	2307	ATACAGATGCAATTTGAAGAATATTTTACTCATGTTTCCAAGAGTGGTGGCTGTGATCAA	2366
DB	1561	ATACAGATGCAATTTGAAGAATATTTTACTCATGTTTCCAAGAGTGGTGGCTGTGATCAA	1620
QY	2367	GAACCTTTACTTGGGCTTATAAAGATGTCAACAAAGCTGTGATGAGGTGACAGTACCAGT	2426
DB	1621	GAACCTTTACTTGGGCTTATAAAGATGTCAACAAAGCTGTGATGAGGTGACAGTACCAGT	1680
QY	2427	TTCATATCTAGTACAGACAGTAGTACAAATCGTGTGGACATAGTTTGGAAACAAGTCC	2486
DB	1681	TTCATATCTAGTACAGACAGTAGTACAAATCGTGTGGACATAGTTTGGAAACAAGTCC	1740
QY	2487	TACAGAGTATCTGAGGATCTTGTAAAGCATACATCTGCCACTCTCTAGGACCCCTTGCTGGT	2546
DB	1741	TACAGAGTATCTGAGGATCTTGTAAAGCATACATCTGCCACTCTCTAGGACCCCTTGCTGGT	1800
QY	2547	CTTCATGTACCTTTTAAAGCAGGCTGGGTCTCTTTCAAGACTGCATGAATTTGTGCTTTT	2606
DB	1801	CTTCATGTACCTTTTAAAGCAGGCTGGGTCTCTTTCAAGACTGCATGAATTTGTGCTTTT	1860
QY	2607	GAGGACTTTCAAGTAGAGGTACTAGTGGAAATATCCTTTTACGTTGTCTGGTGTGGTGGCC	2666
DB	1861	GAGGACTTTCAAGTAGAGGTACTAGTGGAAATATCCTTTTACGTTGTCTGGTGTGGTGGCC	1920
QY	2667	CAGGTTGTCTGAGATGTGCGGAAGAAATGGACTGCTCTTATATAGCCAGGCTTTTAT	2726
DB	1921	CAGGTTGTCTGAGATGTGCGGAAGAAATGGACTGCTCTTATATAGCCAGGCTTTTAT	1980
QY	2727	TACCAAGATGTTAAGTCAGAGAAGAAATGATGATAAAGATATCATCATGCTTCAGATT	2786
DB	1981	TACCAAGATGTTAAGTCAGAGAAGAAATGATGATAAAGATATCATCATGCTTCAGATT	2040
QY	2787	GGTGCATCTTTAATGGATCCCAATTAAGTCTTGTGTACTGTTGTTGAGGTTATGAACCT	2846
DB	2041	GGTGCATCTTTAATGGATCCCAATTAAGTCTTGTGTACTGTTGTTGAGGTTATGAACCT	2100
QY	2847	CCCGAGGCTTTTACAGACCATATCTACAAAGACACAGGATTTGATTAACAATATATAT	2906
DB	2101	CCCGAGGCTTTTACAGACCATATCTACAAAGACACAGGATTTGATTAACAATATATAT	2160
QY	2907	ACACTAATAGAGAATGCTTCAGTCTCTCATCTATATTTGGTGGAGGTTATGACCT	2966
DB	2161	ACACTAATAGAGAATGCTTCAGTCTCTCATCTATATTTGGTGGAGGTTATGACCT	2220
QY	2967	GGAGTGGGAATGTGACCAAGAAGAGGTCACAATGAGAGAAATCATTCACCTTGCTTTGC	3026
DB	2221	GGAGTGGGAATGTGACCAAGAAGAGGTCACAATGAGAGAAATCATTCACCTTGCTTTGC	2280
QY	3027	ATTGAACCCATGCCACAGTGCATTTGCCAAAATTTTACCTGAGAATGAATAATGAA	3086
DB	2281	ATTGAACCCATGCCACAGTGCATTTGCCAAAATTTTACCTGAGAATGAATAATGAA	2340
QY	3087	ACTGGCTTAGAGAATGTCATAAACAAGTGGCCACATTTAAGAACCCAGGTGTATCAGGC	3146
DB	2341	ACTGGCTTAGAGAATGTCATAAACAAGTGGCCACATTTAAGAACCCAGGTGTATCAGGC	2400
QY	3147	CATGGAGTTTATGAACATAAAGATGAATCACTGAAAGACTTCAATATGCTTTTATCAT	3206
DB	2401	CATGGAGTTTATGAACATAAAGATGAATCACTGAAAGACTTCAATATGCTTTTATCAT	2460
QY	3207	TACTCCAAAACCCAGCATAGCAAGGCTGAACATATGCAAGAGAAAAGGAGAAAACAAGAA	3266
DB	2461	TACTCCAAAACCCAGCATAGCAAGGCTGAACATATGCAAGAGAAAAGGAGAAAACAAGAA	2520
QY	3267	AACAAAGATGAAGATTTGCCGCCACCCACCTCTCTGAAATTTCTGCCCTTTCAGCAAA	3326
DB	2521	AACAAAGATGAAGATTTGCCGCCACCCACCTCTCTGAAATTTCTGCCCTTTCAGCAAA	2580

QY	3327	GTGATTAACTTCTCAACTGTGATATCATATGATGTACATTTCTCAGGACCGTATTTGACGG	3386
DB	2581	GTGATTAACTTCTCAACTGTGATATCATATGATGTACATTTCTCAGGACCGTATTTGACGG	2640
QY	3387	GCAATAGACACAGATTCTAACTTGTGGACCGAAGGATGCTCCAAATGGCTTTTCATATT	3446
DB	2641	GCAATAGACACAGATTCTAACTTGTGGACCGAAGGATGCTCCAAATGGCTTTTCATATT	2700
QY	3447	CTGGCATTTGGGTTTACTAGAAGAGAGCAACAGCTTCAAAAAGCTCCTGAAGAAGAGTA	3506
DB	2701	CTGGCATTTGGGTTTACTAGAAGAGAGCAACAGCTTCAAAAAGCTCCTGAAGAAGAGTA	2760
QY	3507	ACATTTGACTTTTATCATATAGGCTTCAAGATTTGGGAAGTTTCAAGCATGATATCAAAATG	3566
DB	2761	ACATTTGACTTTTATCATATAGGCTTCAAGATTTGGGAAGTTTCAAGCATGATATCAAAATG	2820
QY	3567	CTTTTGGAAAAACTCAAAGGAATTTCCCAAGTTAGAAAGGCCAGAAAGACATGATAACCTGG	3626
DB	2821	CTTTTGGAAAAACTCAAAGGAATTTCCCAAGTTAGAAAGGCCAGAAAGACATGATAACCTGG	2880
QY	3627	ATACTTCAGATGTTTGACACAGTGAAGCGATTAAAGAGAAAAATCTTTTAAATTTAGCA	3686
DB	2881	ATACTTCAGATGTTTGACACAGTGAAGCGATTAAAGAGAAAAATCTTTTAAATTTAGCA	2940
QY	3687	ACCACATCAGGATCGGAATCTATTAGAATGATGAGATTACTCATGATAAAGAAAAAGCA	3746
DB	2941	ACCACATCAGGATCGGAATCTATTAGAATGATGAGATTACTCATGATAAAGAAAAAGCA	3000
QY	3747	GAACGAAAAAGAAAAGCTCAAGCTGCTAGGCTACATCGCCAGAAAGATCATGGCTCAGATG	3806
DB	3001	GAACGAAAAAGAAAAGCTCAAGCTGCTAGGCTACATCGCCAGAAAGATCATGGCTCAGATG	3060
QY	3807	TCCTGCTTACAGAAAAAATCTTCAATGAAACTCATATAAACTCATGATGATGACAAATACATCAGAA	3866
DB	3061	TCCTGCTTACAGAAAAAATCTTCAATGAAACTCATATAAACTCATGATGATGACAAATACATCAGAA	3120
QY	3867	ATGCTTGGGAAAGAAAGATTCCCATTTAGGAGGAAGAGAGACCCAGAGTCACTGACTAC	3926
DB	3121	ATGCTTGGGAAAGAAAGATTCCCATTTAGGAGGAAGAGAGACCCAGAGTCACTGACTAC	3180
QY	3927	TCAGAAATTTGCTTTGGGTCTTAAACGGGTCCATCTGTTACTGAAAGAGGTTGCTGACG	3986
DB	3181	TCAGAAATTTGCTTTGGGTCTTAAACGGGTCCATCTGTTACTGAAAGAGGTTGCTGACG	3240
QY	3987	TGCATCTTTGCCAAGAAGAACAGGAGTGAATAATGAAATAATGCGCATGTTATTATCG	4046
DB	3241	TGCATCTTTGCCAAGAAGAACAGGAGTGAATAATGAAATAATGCGCATGTTATTATCG	3300
QY	4047	GCCTGTGTCAGAAATCTACTGCCCTTAACCCAGCACAGGGGAAAAACCCATAGAACTCTCA	4106
DB	3301	GCCTGTGTCAGAAATCTACTGCCCTTAACCCAGCACAGGGGAAAAACCCATAGAACTCTCA	3360
QY	4107	GGAGAAAGCCCTAGACCCACTTTTCATGGATCCAGACTTGGCATATGGAACCTTATACAGGA	4166
DB	3361	GGAGAAAGCCCTAGACCCACTTTTCATGGATCCAGACTTGGCATATGGAACCTTATACAGGA	3420
QY	4167	AGCTGTGTCATGTAATGACGACGATGCTGCTGCGAGAAGTATTTTGAAGCTGTACAGCTG	4226
DB	3421	AGCTGTGTCATGTAATGACGACGATGCTGCTGCGAGAAGTATTTTGAAGCTGTACAGCTG	3480
QY	4227	AGCTCTCAGCAGCGCATTCATGTTGACCTTTTGGAAAGTGGAGAAATATCTTTGC	4286
DB	3481	AGCTCTCAGCAGCGCATTCATGTTGACCTTTTGGAAAGTGGAGAAATATCTTTGC	3540
QY	4287	CCTCTTTGCAATCTCTGTGCAATACTGTATCCCATATTTCTTCTTGGAAACCTCAAAAG	4346
DB	3541	CCTCTTTGCAATCTCTGTGCAATACTGTATCCCATATTTCTTCTTGGAAACCTCAAAAG	3600
QY	4347	ATAAAGACGTGAGAAATCCAGATGCTCTGCTCAACTTTTGGACCTGGCACCGTGGATACAG	4406
DB	3601	ATAAAGACGTGAGAAATCCAGATGCTCTGCTCAACTTTTGGACCTGGCACCGTGGATACAG	3660
QY	4407	ACTGTTCTGGCCAGAAATATCAGGTTTATATATAAGACATGCTAAAGAGGAGAAAAACCCCAATT	4466

308	GTGTACAGATGTC	CAATAC	TACTCCTTT	GGAGTGGTACTTATTT	GGAGAGGATGCCGATA	367
949	TTTTCCTTAGAGAAAT	TGAACACAGCAGTGGACATTT	CAGCTTTT	GGGAGGGTTT	CAAAA	1008
368	TTTTCCTTAGAGAAAT	TAAACACAGTGGAGCGTT	CCAGTTGT	GGGAGGTTT	CAAAA	427
1009	GTGGAGAGACAACCT	ATATCTTTG	CAGGATTTG	CAATATGAT	CAACAATGTTACTCTGTA	1068
428	GTGGAGAAAACAACAT	ATCTCTGT	AGGATTTG	CAATTTGAT	CAACGTTGCTCTGTA	487
1069	TGGACTGCTTCCAGGACAGT	TTTCAT	AAAAATCATCGTT	TACAGATGCATCT	TCTACTG	1128
488	TGGACTGCTTCCAAAGTAGT	TTTCAT	AAAAACCATCGTT	TACAGATGCATCT	TCTACTG	547
1129	GAGGAGGGTTCTGTGACT	GTGGAGACACAGAGCAT	TGGAAACCT	TGGCCCTTTT	TGTGTAA	1188
548	GAGGGGCTTCTGTGACT	GTGGAGACACAGAGCGT	TGGAACCT	TGGCCCTTTT	TGTGTGG	607
1189	ATCATGAACCTGGAAGACAGT	ACTAT	AAAAAGAAATTCACGCT	GTCCGTT	TGAATGAAG	1248
608	ATCACGAGCCTGGAAGACAGT	ACTAT	AAAAAGAGAGCTT	TACAT	TGCCATGAATGAAG	667
1249	AGGTAAATGCTCCAAAGCCAGAGAAAT	ATATTTCC	TTTCAGTGAT	AAAAATATGTCGT	TGAGAAATGA	1308
668	AGGTGAATGCTCAAGCCAGAGAAAT	ATTTCC	TTTGGTGAAT	AAAAATATGTCGT	TGAGAAATGA	727
1309	CTATATGGGAAGAGGAAAAGAACT	CGCTCCT	TGAACTCCAGAT	AAAGGAGAAAATGAAA	1368	
728	CTATATGGGAAGAGGAAAAGAAAT	TGCTCCT	TGAACTTCAGAT	AAAGGAGAAAATGAAAC	787	
1369	GATACATATTTGCTCTTTTCAAT	GATGAACACCATTCAT	TATGAC	CACGTCATATACAGCC	1428	
788	GATACATATTTGCTCTTTTCAACGAT	GAGCACCATTCGAT	TATGATCATGTCAT	CTACAGTC	847	
1429	TACAAGAGCTCTTGACTGT	GAGCTGCAGAGCGCCAGTTG	CAATAC	CACCTGTCATGACA	1488	
848	TGCAGAGAGCTTAGATTTG	CGCTTGCAGAGCACAGCT	TGCACAGCT	TGCCATCGACA	907	
1489	AAGAGGCTCGCGGCTGTTAAAG	CGGAGCTTATGCTGCTGCCAGG	AGCAAGGAAG	1548		
908	AAGAGGCTCGCGGCTGTTAAAG	CGGAGCTTATGCTGCTGCCAGG	AGCAAGGAAG	967		
1549	ATATAAGAGTCAATTCAGAAAAT	GTCTCTCAACATCCACATTCAT	GTAGAAGTAT	TACACT	1608	
968	ATATAAGAGTCAATTCAGAAAAT	GTCTCTCAACATCCACATTCAT	GTAGAAGTAT	TACACT	1027	
1509	CAGAGATATGGCTCATCAGAAA	ATTTGCTTTGGCTTTGGTCT	TGGTCTCCTGGAT	GAACAAAATTA	1668	
1028	CCGTGGTTATGGCTCACAGAAA	ATTCGCTCTGGGCTTTGGCT	TGGCTCTCCTGGAT	GAACAAAATTA	1087	
1669	TGAGCTATTTCAAGTGACTTT	TAGCAGATCTTTTGGCCAAAGCAT	TGCCATG	CCCTTAGAGAGAACCCTG	1728	
1088	TGAGCTATTTCAAGTGACTTT	TAGCAGATCTTTTGGCCAAAGCAT	TGCCATG	CCCTTAGAGAGAACCCTG	1147	
1729	ACTCGGAGATCCCTGTCTCAT	AGCAGGTTAATGCTTTGGGAT	GCAAAAGCTTT	TATAAAG	1788	
1148	GCCTGAAATCCCTGTCTTAT	AGCAGACTAATGCTTTGGGAT	GCAAAAGCTTT	TATAAAG	1207	
1789	GTGCCCCGTAAGATCCCTT	CATGAATGATCTT	CAGCAGTTTTTTT	TATGGAGATGGAATACA	1848	
1208	GTGCCCCGTAAGATCCCTT	CATGAATGATCTT	AGTACTTTT	TATGGAGATGGAATACA	1267	
1849	AAAACTCTTTTGTATGGAAT	TTGTGAAGTATTTATAA	CAACACTGCAGAAA	AGAAATATACA	1908	
1268	AAAACTCTTTTGTATGGAAT	TTGTGAAGTATTTATAA	CAACACTGCAGAAA	AGAAATATACA	1327	
1909	GTGATGATGATGACAGAGT	ATCTCTATAACTGCATTT	CAGTTCAGATGTT	TCTACTGTT	1968	
1328	CGCAGCACCAGAGAGAGT	ATCTCTATAACTGCATTT	CAGTTCAGATGTT	TCTACTGTT	1387	
1969	CTACTCTGGCTCGACATCT	TATTAAGAGCAGAAAT	GTATCTGCT	CAATTAAGAACTC	2028	
1388	CGACCTTGGCCGSCATCT	TATTAAGAGCAGAAAT	GTATCTGCT	CAATTAAGAACTC	1447	

Qy	3109	ACAAAGTGGCCACATTTAAGAAACCGAGGTGTATCAGGCCATGAGTGTATATGAACATAAAAG	3167
Db	2528	ACAAAGTGGCCACATTTAAGAAACCGAGGTGTGCGGCCATGGAGTTATATGAATTTGAAG	2587
Qy	3169	ATGAATCACATGAAGACTTCAATATGTACATTTTATCATTTACTCCAAACCCACCATAGCA	3228
Db	2588	ATGAATCACATGAAGACTTCAATATGTACATTTTATCCATTTATTTAAACACACATAGCA	2647
Qy	3229	AGGCTGAACATATGCAGAGAAAGGAGAAAAACAAGAAAAAGATGAAGCATTTGCCGC	3288
Db	2648	AGGCTGAACATATGCAGAGAAAGGAGAAAAACAAGAAATAAGATGAAGCATTTGCCGC	2707
Qy	3289	CACACACACCTCCTGAAATTCCTGCCCTCTTCAGCAAAAGTGATTAACCTTTCTCAACTGTG	3348
Db	2708	CGCCACCTCTCCAGAGTTCTGCCCTGCTTTACAGAAAGTAGTCAACTGCTCAGCTGTG	2767
Qy	3349	ATATCATGATGTACATTTCTCAGGACCGCTATTTTCAGCGGGCAATAGACACAGATTTCTAACT	3408
Db	2768	ATGTTATGATATACATCTCAGGACCATCTTTTCAGCGGGCAGTGGACACGGAGTCTAACT	2827
Qy	3409	TGTGGACCGAAGGAGTCTCAATGGCTTTTCATATTTCTGGCATTTGGGTTTACTAGAAG	3468
Db	2828	TGTGGACAGAAGGAGTCTCGCATGGCTTCCATATATTTGGCACTTTGGCTTCTCTGGAAG	2887
Qy	3469	AGAAGCAACAGCTTCAAAAAGCTCTCTGAAGAAAGTAAATTTTGACTTTTATCATAAAGG	3528
Db	2888	AGAAGCAGCAGCTTCAGAAAAGCTCTCAAGAGGAAGTGGCTTTTGACTTTTACCATAAAG	2947
Qy	3529	CTTCAAGATTTGGGAAGTTCAAGCCAT-----GAATATACAAATGGCTTTTGGAAAAAC	3579
Db	2948	CTTCAAGATTTGGGAAGTTCAAGCCATGAATGCTCAGAAATATACAAATGGCTTTTGGAAAGAC	3007
Qy	3580	TCAAAGGAATTCCTCCAGTTTGAAGGGCCAGAAAGACATGATAAGCTGGATACTTTTCAGATGT	3639
Db	3008	TCAAAGGAATTCCTCCAGTTTGAAGGGCCAGAAAGACATGATAAGCTGGATACTTTTCAGATGT	3067
Qy	3640	TTGACACAGTGAAGCGATTAAGAGAAAAATCTTTGTTTAAATGTAGCAACACATCAGGAT	3699
Db	3068	TTGACACAGTGAAGCGATTAAGAGAAAAATCTTTGTTTAAATGTAGTGTGGCAACCATTCAGGAC	3127
Qy	3700	CGGAATCTATTAGAATGATGAGATTACTCATGATAAGAAAAAGGCAGACGAAAAAGAA	3759
Db	3128	TGGAGTGCAATTAAGAGTGAAGGATTAATCTCATATAAGAAAAAGGCAGACGGAAGAGAA	3187
Qy	3760	AAGCTGAAGCTGTAGGTCATCTGCAGAAAGATCATGGCTCAGATGTCTGTCCTTACAGA	3819
Db	3188	AAGCTGAAGCGCTAGGCTTCATGCCAGAAAGATCATGGCCAGATGTCTGCCCTTACAGA	3247
Qy	3820	AAAACCTTATTTGAAACTCATAACTCATGTATGACAAATACATCAGAAATGCCTGGGAAAG	3879
Db	3248	AAAACCTTATTTGAAACCCACAACTCATGTATGATAATACGTACAGAAGTAACAGGGAAGG	3307
Qy	3880	AAGATTTCCATATTGGAGGAAGACAGACACCCAGCCAGTCAGTGACTCTAGAAATTTGCTT	3939
Db	3308	AAGACTCCATATTGGAGGAAGACAGACACCTTCAGCAGTCAGTGAGGCTCTAGAAATTTGCTC	3367
Qy	3940	TGGTCTCTAAACGGGGTCCATCTGTTACTGAAAAGGAGGTGCTGAGTGCATCCTTTGCC	3999
Db	3368	TGGGCCCTTAACGGGGCCCGCTGTTTACCGAAAAGGAGGTGCTGACGTGCATCCTCTGCC	3427
Qy	4000	AAGAAGAACAGGAGGTGAAAATAGAAAAATATGCCATGGTATTATCGGCTGTGTCCAGA	4059
Db	3428	AAGAAGAACAGAGGTAAAACTAGAAAAATATGCCATGGTATTGTACAGCATGTGTGCAGA	3487
Qy	4060	AACTACTTGCCTTAAACCCACACAGGGGAAAAACCCATAGAACTCTCAGGAGAAAGCCCTAG	4119
Db	3488	AATCCACCGCCCTAAACCCACACAGAGGGGAAGGCTGTGGACCACTTATAGGGGAAACACTGG	3547
Qy	4120	ACCCATTTTTCATGATTCAGACTTTGGCATATGGAATTTATACAGGAAGCTGTGGTCAATG	4179
Db	3548	ACCTCTTTTCATGGATTCAGACTTTGGCACATGGAATTTATACAGGAAGCTGTGGTCAATG	3607
Qy	4180	TAATGCAAGCAGTGTGCTGGCAGAAGTATTTTGAAGCTGTACAGCTGTACAGCTCTCAGCAGC	4239

[illegible]

QY	1309	CTATATGGGAAGAGGAAAAAGAACTGCCTCCTGAACCTCCAGATAAGGGAGAAAAAATGAAA	1368
DB	728	CTATATGGGAAGAGGAAAAAGAAATGGCTCCTGAACTGCAGATAAGGGAGAAAAAATGAAC	787
QY	1369	GATACTATTGTGTGCTTTTCAATGATGAACACCAATTCATATGACCACGTCATATACAGCC	1428
DB	788	GATACTATTGTGTGCTTTTCAAGCATGAGCACCAATTCGTATGATCATGTGATCTACAGTC	847
QY	1429	TACAAAGAGCTCTGACTGTGAGCTCGCAGAGCCCCAGTTGCGATACCACATGCCATTTGACA	1488
DB	848	TGCAGAGAGCTCTAGATTGGAGGCTTCGACAGGCACAGCTGCACAGACTGCCATCGACA	907
QY	1489	AAGAGGCTCGTCGGGCTGTTTAAAGCGGGAGCTTATGCTGCTGCAGAGAGCAAAAGGAAG	1548
DB	908	AAGAGGCTCGCGGGCTGTCAAGCAGAGTGTGATGACCATTGCCAGGAAGCAAGAGGAGG	967
QY	1549	ATATAAGAGTCATTTCAAGAAATGTCTCTCAACATCCACTTCATGTAGAAGTATTACACT	1608
DB	968	ATATAAGAGTCACTCAGAGAACGTCTCTCAGCACCCCTCCATGTGGAAGTGTCTGCACT	1027
QY	1609	CAGAGATTATGGCTCATCAGAAATTTGCTTTTGGTCTTGTGATGAACAAAAATTA	1668
DB	1028	CCGTGGTATTGGCTCACAGAAATTCGCTCTGGGCCCTGGCTCTGGATGAACAAAAATTA	1087
QY	1669	TGAGCTATTCAAGTGACTTTAGGCGAGATCTTTTGCCAAAGCATGCCTTAGAGAGAAGACCTG	1728
DB	1088	TGAGCTATTCAAGTGACTTTAGACAGATATTTTGCCAGGCTCGCTCGTAGAAGACCTG	1147
QY	1729	ACTCGGAGATCCCTGTCTCATAGCAGGTTAATGCTTTGGGATGCAAGCTTTTATAAG	1788
DB	1148	GCTCTGAANAATCCCTGTCTTATAGCAGACTAATGCTTTGGGATGCAAACTTTATAAG	1207
QY	1789	GTGCCCCGTAGATCCTTCATGAATGTGATGATTCAGCAGTTTTTTTTATGGAGATGGAATACA	1848
DB	1208	GTGCCCCTAAGATCCTTCATGAATGTGATCTTACTAGTTTTTTTATGGAGATGGAATACA	1267
QY	1849	AAAAACTTTTGCTATGGAATTTGTGAAGTATTATAAACAACTGCAGAAAGAATATATCA	1908
DB	1268	AAAAACTTTTGCTATGGAATTTGTGAAGTATTATAAACAACTGCAGAAAGAATATATCA	1327
QY	1909	GTGATGATCATGACAGAAGTATCTCTATACTGCACPTTCAGTTTCAGATGTTTACTGTTC	1968
DB	1328	CGACGACGACGAGAGAAGCATCTCCATAACGCCCTGTCCGTGCGAGATGCTCACCGTCC	1387
QY	1969	CTACTCTGGCTCGACATCTTATTGAAGACGAGAAATGTTATCTCTGTCATTACTGAAACTC	2028
DB	1388	CGACCTTGGCCCGGCATCTTATTGAAGACGAGAAATGTTATCTCTGTCATTACTGAAAGC	1447
QY	2029	TGCTAGAAGTTTACCTGTAGTACTTGGACAGGAACAATAAATTCACCTCCAGGGTTATA	2088
DB	1448	TGCTAGAAGTTTACCTGGAATACTTGGACAGGAACAATAAATTCACCTCCAGGGTTATA	1507
QY	2089	GCCAGGACAATGGGAAGAGTATATCGACTAATATGTCACCTTAAGATATATCTGATCA	2148
DB	1508	GCCAGGACAACGGGAAGAGTCTACGCACTTATGTGACCTTAAAGTATATCTGATTA	1567
QY	2149	GCAAAACCAATATTGGACAGAAAGATTAAGAATGCAAGTTCCTTGAAGGTTTTTCGATCTT	2208
DB	1568	GCAAGCCTGTATATGGACAGACGATTAAGAGGCGAGTTCCCTGGGAAGGTTTCGGGTCTT	1627
QY	2209	TTTTGAAGATTCTTACCTGTATCGAGGGAATGGAAGAAATCCGAAGACAGGTTTGGGCAAC	2268
DB	1628	TTCTGAAGATTCTTACCTGTATCGAGGGAATGGAAGAAATCAGAGACAAGTTGGACAAC	1687
QY	2269	ACATTGAAGTGGATTCCTGATTGGGAGGCTGCCAATGCTATACAGATGCAATTAAGAATA	2328
DB	1688	ACATTGAAGTGGACCCCTGACTGGGAGGCTGCCATCGCTATACAGATGCAACTTAAAGAATA	1747
QY	2329	TTTTTACTCATGTTCCAAAGATGGTGTGCTTTGTGATGAAGAACTTCTACTTGTGGCTTATA	2388
DB	1748	TTTTTGTCTATTCTCCRAAGTGGTGTGCTTTGTGATGAAGATCTTCTACTGGTGGCTTATA	1807

QY	2399	AAGAAATGTCACAAAGCTGTGTAAGGTGCAGTACCACTTTCATATCTAGTAGCAAGACAG	2448
DB	1808	AAGAAATGTCACAAAGCTGTAATGAGGTGCAGTACAAATCTATCTCCAGTACCAAGACAG	1867
QY	2449	TAGTACAAATCGTGTGGACATAGTTTGGAAACAAAGTCCCTACAGAGTATCTCAGGATCTTG	2508
DB	1868	TAGTGCAAATCTGCGGTCTATAGTCTGGAACCAAAATCCTACAAAGTGTCTGAGGACCTTG	1927
QY	2509	TAAGCATACATCTGCCACTCTCTAGGACCCCTGCTGTCTTCATGTACGTTTAAAGCAGGC	2568
DB	1928	TAAGCATACACTGCCACTCTCTAGAACACTTGTGTGTCTTCATGTACGTTTAAAGCAGAC	1987
QY	2569	TGGTGTCTGTTTCAAGACTGCAATTTGTGTCTTTTGAGGACTTTTCAAGTAGAGGTAC	2628
DB	1988	TAGTGTCTATTTCAAGACTGCATGAATTTGTGTCTTTTGACAGCTTTCAAGTAGAGGTCC	2047
QY	2629	TAGTGCGAATATCCTTTAGCTTGTCTGTGTGTGTGTGCCCAGGTGTGTCTGAGATGTGCC	2688
DB	2048	TGTTGTGAGATACCCCGCTGCGCTGCTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	2107
QY	2689	GAAGAAATGGACTCTCTCTTATTAGCCAGGTGTTTTATTACCAAGATGTTTAAAGTGCAGAG	2748
DB	2108	GAAGAAAGGGCTCTCACTCATCAGCCAGGTTTTCTATTATCAAGATGTTTAAATGCAGGG	2167
QY	2749	AAGAAATGTATGATAAAGATATCATCATGTCTCAGATTTGGTGTGATCTTTTAAATGGATCCCA	2808
DB	2168	AGGAAATGTACGATAAAGATATCATCATGCTTCAGATTTGGAGCATCTATTAATGGATCCCA	2227
QY	2809	ATAAGTTCTTGTACTGGTACTTCAGAGGTATGAACTTGTCCGAGGCTTTTAAACAAGNCCA	2868
DB	2228	ACAAGTTCTTGTACTGGTACTTCAGAGATATGAACTTACTGTACTGTCTTTTAAACAAGACCA	2287
QY	2869	TATCTACAAAGACCCAGGATTTGATTTAAACAAATATAACACTTAATAAGAAAGAAATGCTTC	2928
DB	2288	TATCCACAAGACCCAGGATTTGATTTAAACAGTATATACATTAATAGAGAGAAATGCTTC	2347
QY	2929	AGGTCTCATCTATATTGTGGTGAGCGTTTATGTACCTGGAGTGGGAAATGTGACCAAG	2988
DB	2348	AGGTCTCATCTATATTGTGGAGAACGTTATGTACCTGGAGTGGGAAATGTTACCAGAG	2407
QY	2989	AAGAGGTCACAATCAGAGAAATCATTCAC TTGCTTTGCATTTGAACCATGCCACACAGTG	3048
DB	2408	AGSAGGTTATATCAGAGAGATTACTCAC TTACTTTTGCATTTGAGCCCATGCCACACAGTG	2467
QY	3049	CCATTCGCAAAATTTACCTTGAGAAATGAAATATGAAACTGGCTTAGAGAAATGTCATAA	3108
DB	2468	CCATTCGCCAGNAACCTACCTTGAGAACGAAATATGAAACTGGCTTAGAGAAATGTCATAA	2527
QY	3109	ACAAAGTGGCCACATTTAAGAAACCCAGGTGTATCAGGCCATGGAGTTTATGAACATAAAG	3168
DB	2528	ACAAAGTGGCCACATTTAAGAAACCCAGGTGTGTCGGGCCATGGAGTTTATGAATGGAAG	2587
QY	3169	ATGAATCACTCAAGAGCTCAATATGTACTTTTATCATTTACTCCAAACCCACGATAGCA	3228
DB	2588	ATGAATCACTGAAAGAGCTCAATATGTACTTTTACCATTATCTTAAACACACGATAGCA	2647
QY	3229	AGGCTGAACATATTCGCAAGAAAAAGGAGAAAAACAAGAAAAACAAGATGAAGCATTTGCCG	3288
DB	2648	AGGCTGAACATATTCGCAAGAAAAAGGAGAAAAACAAGAAAAACAAGATGAAGCATTTGCCG	2707
QY	3289	CACACCAACCTCTGAATCTCTGCCCTGCTTTCAGCAAAAGTATTAACCTTCTCAACTGTG	3348
DB	2708	CGCCACCTCTCCAGAGTTCTGCCCTGCTTTTCAGCAAAAGTAGTCAACCTGCTCAGCTGTG	2767
QY	3349	ATATCATGTATGATACATTTCTCAGGACCGTATTTGAGCGGGCAATAGACACAGATTTCAACT	3408
DB	2768	ATGTTATGATATACATCTCAGGACCATCTTTGAGCGGGCAGTGGACACAGGAGTCTTAATC	2827
QY	3409	TGTGACCCGAAGGAGTCTCAATGCTTTTTCATATTTCTGGCATTTGGGTTTTACTAGAAG	3468
DB	2828	TGTGGACAGAAAGGATGCTCAGATGGCGTTCCATATATTGGCACTGGGCTTCTCTGGAG	2887
QY	3469	AGAAGCAACAGCTTTCAAAAAGCTCCTCGAAGAAAGAGTAACTTTTATCTTTTATCATAGG	3528

Db	2888	AGAGCAGCAGCTTCAGAAAGCTCCTGAAGAGGAGTGGCTTTTGACATTTTACCATAAAG	2947	Db	3968	AGTCTTGGGTGAAATATTCAAATAGTATCAAGGAAATGGTCAATTCTCTTCGCCACAACAA	4027
Qy	3529	CTTCAAGATTGGGAAGTTCACCCAT-----GAATATACAATGCTTTTGGAAAAAC	3579	Qy	4597	TTTATAGAAATTGGATTGAAAGTGCCACCTGATGAAAGGATCCTCAGATCCCATGCTGA	4656
Db	2948	CTTCAAGATTGGGAAGTTCACCCATGMAATGCTCAGAAATATACAATGCTTCTTGGAAAGAC	3007	Db	4028	TTTACAGAAATTGGCCTGAAAGTGCCCTGATGAACATAGACCACGAGTGGCCCATGATGA	4087
Qy	3580	TCAAGGAATTCGCCAGTTAAGAGCCGAGAGGACATGATACGTTGGATCTTCCAGATGT	3639	Qy	4657	CCTGGAGCAGCTTGGCTTTTCACTATCCAGGCAATTCGAAATCTATTGGGAGATGAAGGAA	4716
Db	3008	TCAAGGAATTCGCCAATTAGAAGCCGAGAGGACATGATACATGGATCTTCCAGATGT	3067	Db	4088	CCTGGAGCAGCTTGGCTTCCATCCAGGCAATCGAAACCTTGTGGGAGATGAAGGAA	4147
Qy	3640	TTGACACAGTGAAGCGATTAGAGAGAAATCTTGTAAATGTAGCAACACCATCAGGAT	3699	Qy	4717	AACCTCTGTTTGGAGCAGCTTCAAAATAGGCAGCATTAATGGTCTGAAGCAATTAATCGAT	4776
Db	3068	TTGACACAGTGAAGCGATTAGAGAGAAATCTTGTAGTTGTGCAACCATCTCAGGAC	3127	Db	4148	AACCTCTAATTGGAGCAGCTTCAAAATAGACAGCATAGCGTCTGAAGCGCTAATCGAGT	4207
Qy	3700	CGGAATCTATTAGAATGATGAGATTACTCATGATATAAGAAAAAGCAGCAAGCAAAAGAA	3759	Qy	4777	TTGCAAGTGCACAGAGGATTAACCTGTCTCAGGCTCTGTATACAGAAACATCTGTGGTCTC	4836
Db	3128	TGGAGTGCATTTAAGAGTGAGGAGATTACTCATGATATAAGAAAAAGCAGCAAGCAAGAA	3187	Db	4208	TTGCAAGTGCACAGAGGCTACCTGGCCTCAGGCTCTGTATACAGAAACATCTGTCTCGGC	4267
Qy	3760	AAGCTGAAGCTGTAGGCTACATCGCCAGAGATCATGGCTCAGATGTCTGCCTTACAGA	3819	Qy	4837	TTCTATCAGTTGTCTTCTTAACATAAAATCAGAAATACACCATGCTTCTCTCTATAG	4896
Db	3188	AAGCTGAAGCGCTAGGCTTCATCGCCAGAGATCATGGCCAGATGTCTGCCTTACAGA	3247	Db	4268	TCCCTGTCAGTTATCTTCTTAACCTGCAATCAGAAATACACGAGCCTTCTGTCTGTGG	4327
Qy	3820	AAACTTCAATGAACCTCATAACTCATGTATGACAATACATCAGAAATGCTGGGAAAG	3879	Qy	4897	ATCTGTTTCATGTTTGGTGGGTGCTGTGTAGCATTTCCCATCTTGTATGGGATGACC	4956
Db	3248	AAACTTCAATGAACCCCAAACTCATGTATGATAATACGTCAAGATTAACAGGGAAG	3307	Db	4328	ATCTCTTCATGTTCTGTCTGGCGCAGTCTTAGCGTTCCCATCTTGTATGGGATGACA	4387
Qy	3880	AAGATTCCATTTATGAGGAAGAGAGCACCAGCAGTCAGTCAGTACTAGAAATGCTT	3939	Qy	4957	CTGTTGATCTGAGCCCTTCTTCAGTTAGTTCTTCTTATACACACCTTTATCTCTTCCATT	5016
Db	3308	AAGACTCCATTTATGAGGAAGAGAGCACCCTCAGCAGTCAGTCAGGCCCTTAGAATGCTC	3367	Db	4388	CGTGGATCTGACGCGCTGCCACTTAGTTCTTCATATAACCACTCTATCTCTTCCATC	4447
Qy	3940	TGGTCTCTAAACGGGTCCATCTGTTACTGAAAGGAGGTGCTGACGTGCATCTTTGCC	3999	Qy	5017	TCATCACCATGGCAGACATGCTTCAGATACTACTACTACTACAGTAGACA-----CAG	5064
Db	3368	TGGGCCCTTAAACGGGGCCGCTGTACCGAAAGGAGGTGCTGACGTGCATCTCTGCC	3427	Db	4448	TCATCACCATGGCAGACATGCTTCAGATACTCTTACACAGATACAGATCTGTCTCCAG	4507
Qy	4000	AAGAGAACAGAGGTGAAATAGAAATAATGCCATGATATATCGCCCTGTGTCACAGA	4059	Qy	5065	GCCTACCCCTTGCTCAGGTTCAAGAGACAGTCAAGAGGCTCATTCGGCATCTTCTTTCT	5124
Db	3428	AAGAGAACAGAGGTAAACTAGAAATAATGCCATGATATATCGCCCTGTGTCACAGA	3487	Db	4508	GGCGCGCTTGCTGAGGTTGAAGAGATAGTGAAGGAGCTCCTGTGCATCTGCTTCT	4567
Qy	4060	AATCTACTGCCCTTAAACCAGCACAGGGGAAACCCATAGAATCTTCAGGAGAACCCCTAG	4119	Qy	5125	TTGCAAGAAATTTCTCAATATACAAGTGGCTTCCATTTGGGTGTGATATTCCTGGGTGAT	5184
Db	3488	AATCCACCGCCCTAACCCAGCACAGAGGGAAGCCCTGTGACCACCTTAGGGGAAACACTGG	3547	Db	4568	TTGTGGAAGTTCGCGACACACAGACGGCTCTCAGTGGTGGGTGCTCCCGGCTGTAC	4627
Qy	4120	ACCACATTTTCATGATCCAGCTTGGCATATGGAACATTAACAGAGAGCTGTGTCATG	4179	Qy	5185	TGTGGGTCTCACTGAGAAATGGCATCACCCCTTATCTTCGCTGTGCTGATTTGTTTTCC	5244
Db	3548	ACCCTTTTTCATGATCCAGCTTGGCACATGGAACCTTATACAGAGAACTGTGTCATG	3607	Db	4628	TGTGGGTCTCCCTGAGGAACGGCATCACCCCTTACCTCCGCTGTGCTGCACTGCTTTTCC	4687
Qy	4180	TAATGCAGCAGTGTGTCGAGAGATTTTGAAGCTGTACAGCTGAGCTCTCAGCAGC	4239	Qy	5245	ACTATTTACTTGGGGTAACTCCGCCCTGAGGAAGTGCATACCAATTTCTGCAAGAGAGT	5304
Db	3608	TAATGCATGCACTGTGTCGAGAGATTTTGAAGCTGTGCACTGAGCTCGCAGCAGC	3667	Db	4688	ACTATTTACTTGGAGTAGTCTCCGCCCTGAGAACTGTTTGGCAATTTCTGCTGAGGAGAT	4747
Qy	4240	GCATTCATGTTGACCTTTTTCAGTTGGAAGTGGAGAAATATCTTTGCCCTCTTTTGCAAAT	4299	Qy	5305	ACAGTGCACCTGTAGCTATCTATCTTTACCTACAAATTTGTTCTCTCTCTCCAGGAAT	5364
Db	3668	GCATTCAGTAGACCTGTTGACCTGGAGAGCGGAGTACCTATGCCGCTCTGCAAGT	3727	Db	4748	TCAGTGCACCTGTAGCTATCTATCTTTACCCACAAATTTGTTCTCTCTTCCAGGAAT	4807
Qy	4300	CTCTGTGCAATCTGTGATCCCATTAATCTTTGCAACCTCAAAAGATAAAGAGTGA	4359	Qy	5365	ATTGGATCTGTAAAGGCCCTTGTCTCCAGAGGTGGTGTGAGATCTGCTTACTAACT	5424
Db	3728	CTCTCTGCAACACTGTATCCCATCATCCCTTTGCAGCCGAGAAAGATCAACAGTGA	3787	Db	4808	ATTGGATACCATTAAGGCCCTTACTACAGAGTGGTGTGAGATCTGCTTACTCAAGT	4867
Qy	4360	ATGACAGATGCTTGTCTCAACTTTTGACCTTGGCACGGTGGATACAGACTGTCTGGCCA	4419	Qy	5425	GTTTGAAGCAAAAAACACCGGTGAGTACCCCTACCTAGAAAAAAGATTTGATAGAGC	5484
Db	3788	ATGGGAGGCTCTTGTCTCAACTTTTGACCTTGGCCCGTGGATACAGACTGTCTTGCCA	3847	Db	4868	CTTTGAAGCAGAAAAAGTGTGAGTACCCCTACCTAGAAAAAAGATTTGATAGAGC	4927
Qy	4420	GAATATCAGTTTATATAAGACATGCTAAAGGAGAAAAAC-----CAATCTCTATTCTCT	4476	Qy	5485	TTCTCTGATGACTATAGCTGCTCTCTGAATCAAGCTTCTCAATTTACAGTGGCCCGCTG	5544
Db	3848	GAATATCGGTTTATATAAAGCATGCTAAAGGAGAAAGCCCGACGATCTCTGCTGTGT	3907	Db	4928	TTCTCTGAGACTACAGCTGTCTTAATCAGGCTTCTCACTTTAGTGTCCAGGCTGTG	4987
Qy	4477	TTAATCAAGGAATGGGAGATTCTACTTTGGAGTTCCATTTCCATCTCAGTTTGGGTG	4536	Qy	5545	CAGATGATGAGCGAAAGCATCTCTGTCCCTCTGTCCTTTTCTGTGGGCTATACTATGTTCTC	5604
Db	3908	TTAATCAAGGAATGGGAGTTCAACTTTTGAGTTTCAATTTCCATCTCAGTTTGGAGTTC	3967	Db	4988	CAGATGATGAGCGAAAGCATCTCTGTCTCTCTTTTCTGTGGGCGCATCTGCTGTTCTC	5047
Qy	4537	AGTCTTCCGATTAAATATTCAATAGCATCAAGGAATGGTTATCTCTTTGGCCACAACAA	4596	Qy	5605	AGAACATTTGCTGCCAGGAAATTTGTGAACGGGGAAGAGTTGGAGCTTGCATTTTTCACG	5664
				Db	5048	AGAACATCTGTTGCCAAGAAATAGTGAATGGGGAAGAGTGGAGCGTGGCTTTTTCATG	5107

QY	5665	CACCTTCACCTCTGGAGCGGAGTCTGCATTTTCTCTAAATAATCAGAGATGCCGAGTGCTCC	5724
Db	5108	CGCTTCATTTGGTGTGGAGTCTGCATTTTCTCTAAATAATCAGAGATGCCGAGTGCTCC	5167
QY	5725	TGCTTGAAGGTAAAGCCAGAGGCTGTGCCTTATCCAGCTCCTTACTTGGATGAATATGAG	5784
Db	5168	TGGTGAAGGMAAGCCAGAGGCTGTGCCTTATCCAGCTCCTTACTTGGATGAATATGAG	5227
QY	5785	AAACAGACCTTGGCTCTGAAGAGGGGCAACCCCTTCAATTTATCTGCTGAGCGGTATCGGA	5844
Db	5228	AAACAGACCCAGGCTTAAAGAGAGGAAACCCACTTCAATTTATCTGCGGAGCGGTATCGGA	5287
QY	5845	AGCTCCATTTGGTCTGGCAACACACACTGCATTTATAGAAGAGATTTGCTAGGAGCAAGAGA	5904
Db	5288	AGCTGCATTTGGTCTGGCAACACACTGCATTTATAGAAGAGATTTGCTGCGAGCAGGAGA	5347
QY	5905	CTAATCAGATGTTATTTGGATTCAACTGGCAGTTACTGTGAGCTCCAACCTCTGCCTCAAG	5964
Db	5348	CTAATCAGATGTTATTTGGATTCAACTGGCAGTTACTGTGAGCTCTCAGTTCTGCCTCAAG	5407
QY	5965	ACAATCAACAATCAGCAGCAGTAGTAAAGGCTGATTCAAAATTTATGGAATACTTTCTGAGG	6024
Db	5408	ACAATCAGTGCATCAATAATAAGACTGATCTAAAATTTCTAGAGAACTTTCTGAGG	5467
QY	6025	GCTGGGAAGTATTTGAGGAGTCTTTTGCTCCATGTCCAGGTTCCACTTACATCAATAAAT	6084
Db	5468	ACGGGGGAAGTATTTGAGGAGTCTTTTGATCCATGTCCAGATTCACACACATTAATAAAT	5527
QY	6085	ATTTCTTAATGAGTATTCCTTCAATTAGCAACATATGCTTCACAGGAAAAA-AGGAC	6143
Db	5528	ATTCCTTAATGAATATTCCTTCAATTAGCAACATATGCTTCACAGGAAAAAACAAGAC	5587
QY	6144	ATAGATCAATCTCT-----TTTT	6160
Db	5588	ATAGATTAATGTTTATGTTCTAGACACTAAGAATAAGTCTGTTTATCAACAGTGCTAT	5647
QY	6161	ATGTGCTAGTATTTCCAGGAATTTATCCCTTCATAATTTGCTCATTTTATTTATTT	6220
Db	5648	TTCTGCTAATATTTCCAGAAAACTCTTCCCTTCATACTGCTAGTTCAATTTTCATAT	5707
QY	6221	CATCCACTTGGTAGATGAAGTACGTCACGTCACACAGTTGTAGACATTTATGTGTTGTTAAC	6280
Db	5708	CACCCACCTGGTAAATGAGGTACATTAAGCATTTGTGGACATTTCTCCATCTCGGTAAC	5767
QY	6281	TCCTTCGCAATTTTGTATTGGTGTTT	6307
Db	5768	ATCTCTGCACCTTTGTATTGGTGTTT	5794
RESULT 5			
AF061555		6395 bp mRNA linear ROD 10-JUL-1998	
DEFINITION		Mus musculus ubiquitin-protein ligase E3-alpha (Ubr1) mRNA,	
complete cds			
AF061555			
VERSION		AF061555.1 GI:3170886	
KEYWORDS			
SOURCE		Mus musculus (house mouse)	
ORGANISM			
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
REFERENCE		1 (bases 1 to 6395)	
AUTHORS		Kwon,Y.T., Reliss,Y., Fried,V.A., Hershko,A., Yoon,J.K., Gonda,D.K.,	
		Sangan,P., Copeland,N.G., Jenkins,N.A. and Varshavsky,A.	
TITLE		The mouse and human genes encoding the recognition component of the	
		N-end rule pathway	
JOURNAL		Proc. Natl. Acad. Sci. U.S.A. 95 (14), 7898-7903 (1998)	
MEDLINE		98318583	
PUBMED		9653112	
REFERENCE		2 (bases 1 to 6395)	
AUTHORS		Kwon,Y.T. and Varshavsky,A.	
TITLE		Direct Submission	

JOURNAL	Submitted (24-APR-1998) Division of Biology, 147-75, California		
	Institute of Technology, 1200 E. California Bl., Pasadena, CA		
	91125, USA		
FEATURES			
source	location/Qualifiers		
	1. 6395		
	/organism="Mus musculus"		
	/mol_type="mRNA"		
	/db_xref="taxon:10090"		
	/chromosome="2"		
	/cell_line="MEL-C19"		
gene	1. 6395		
	/gene="Ubr1"		
5'UTR	1..114		
	/gene="Ubr1"		
CDS	115..5388		
	/gene="Ubr1"		
	/note="Ubr1p; recognition component of the N-end rule		
	pathway"		
	/codon_start=1		
	/product="ubiquitin-protein ligase E3-alpha"		
	/protein_id="AAC40165.1"		
	/db_xref="GI:3170887"		
	/translation="MADREMGAERMDVSPBPPLAPORPASWMDQVDFTATFLHLHA		
	OLPELYFAEMDPDLERKEESVQMSILTPLEWYLFEGEDPDICLEKIKSGATFOLGKV		
	FKSGEYISCRDCAIDPTCVLCMDCFQSSVHNHRYKMTSTGGGFCDCGDEAKWTG		
	PFCDVDEPGRAGTTKESLHCPLENEVIAQARRIPSPVVIKVIYEMTWEEKEKLPVLQ		
	IREKERYCYVLFNDEHSHYVSIQALDCELAELAQHTTADKGRRAVAGVY		
	ATCQAEADIKSHSENQHPHVEVLVSHVMAHQFALRLGSMWNKIMSYSDSFRQI		
	FCOACLEVEGSENPCLFSRLMLMDAKLYKGARKILHELIFSPFMEMEYKLLFAMEF		
	VKYYKOLQKEYISDDHRSISITALSVMQVPTLARHLIBEONVISVITETLLLEVL		
	EYLDNRNKNFQGYSDQKLGRIYAVICDLKYLILISKPIWTERLAQLEGRFSFKI		
	LTCMQMEETRRQYQHEVDPDMEAAIAIQMLKILLMFQWACDEDDLLLVAYKE		
	CHKAMVSTNFMSTFVQVLCVGHSLTSKYSKVEDLSVSHLPSRLTAGHLVRLSR		
	LGALSRHEFPVDFSVLEVLPRLCLVLAQVAEMWRNGLSLISQVFFYQDVK		
	CREMYDKDIIMLOIGASIMDPNKFLLVLORYELTDAFNKTIPTKDDOLIKQYNTLI		
	EEMLOVLIYVGRYVPGVGNVTRREVMIRETHLLICEPHMPSIAARNLPENNNET		
	GLENINKVATFRKPGVGHGVYELKSDKDFNMFYHYSHQSHKAMHKKRRKQ		
	ENKDEALPPPPPECFAPFSKVNLLSCDVMYIILRTIFERAQVDESNLWTEGMLQMA		
	FHILALGLEEKQOLQAPAEVEAFDFYHKASRLGSSAMNAQNIOMLLERLGIQLE		
	GOKDMITWLOQFDTVKRLREKSLVATTSGLGCIKSEITHDEKAEKRKKAAAR		
	LHRQKIMAOALOKNFLETHKLMYDNTSEVTKGKEDSIWEEESTSAVSEASRIALGPK		
	RGPATKEVLTCLIQBEQEVKLENNAMVLISACVQKSTALTQHRGKPVDHGETLDP		
	LFMDPDLAHGTYTSCGHVMHAWCQKFEAVOLSQQRHVDLEDSELYCELPCK		
	SLCNTVPIIPQPKINSENAEAQLTLARWTQTLVLRISGNIHKAKEAPAV		
	VLFNOGMDSTFEFHSISFGVSKVSNKSEMVILFATTIYRIGLKVPPEDLDP		
	YPMNTWSTCAFTIOAENLDEGKPLFGALQNRHOSGLKALMQRAVAAORATCPQVLI		
	HKHLARLLSVILPNLQSENTPGLLSVDLPHVLVGLAVLAFPSLYWDYDLPQSPLESS		
	YNHLYLFHLITMAHMLQLLTDTDLSPGPPLAEGEEDSEEARCSAFEEVQSHTDG		
	LTCGAPGWLWLSLRNGITPYLRCAALLFHYLLGVAPPPEELFANSAGEFSALCSYL		
	SLPTNLFLLFOEYWDTRIPLQRCMGDPALLSLKQKSAVVYPRKRNLSLIELPDSY		
	CLLNOASHFRCPRSADDERKHPVLCFLCGAILCSQNICCOEIVNGEEVACVPHALHC		
	GAGVCIFLIRECRVVLVEKGARGCAYPAPYLDEYGETDPGLKRGKNPLHLSRERYKL		
	HLVWQQCHIEIEIARSQETNOMLFGFNQLL"		
	5389..6395		
	/gene="Ubr1"		
	5429..5434		
	/gene="Ubr1"		
3'UTR	1802 a 1376 c 1519 g 1698 t		
polya_site			
BASE COUNT	1802 a 1376 c 1519 g 1698 t		
ORIGIN			
	Query Match 70.3%; Score 4435.4; DB 10; Length 6395;		
	Best Local Similarity 86.3%; Pred. No. 0;		
	Matches 4997; Conservative 0; Mismatches 721; Indels 69; Gaps 6;		
QY	590	TTCAGGGCGCGTCTGTAAGATGTCGCTCCTCTCTC-TCCGACCGGCCACAGGTTTCGCGT	648
Db	8	TTACAGGGCGCGTCTGTAAGATGTCGCTCCTCTCTCCTCGCGCGGCCACAGGTTTCGCGT	67
QY	649	TGCCTCTGCGCGGGGTGCGCACTGCAGGCGTCAGTTTCCCTCAAGATGCGGCGGAGG	708
Db	68	AGCTGCGCGCGGGGTGCGGAAGTTCGCTTTCCCTTTAAGATGCGGCGGAGG	127
QY	709	AGGCTGGAGGTACTGAGAGGATGGAATCAGCGCGGAGTTACCCCGACACCCCTCAGCGTC	768

Db	128	AGATGGACGGCGCGAGAGATGGAAGTACGCCGAGAGCCTCCCTTGCCCCGACGCGC	187
Qy	769	TGGCATCTTGGTGGGATCACAAGTTGATTTTATCTACTGCTTCTTCTTGTCATCATTTGGSCAC	828
Db	188	CGGCATCGTGGTGGGATCACAAGTTGATTTCTATCTACTGCTTCTTCTTACATCATTTGGCAC	247
Qy	829	AATTTGGTCCAGAAATTTACTTTTCTGTAATGACCCAGACTTTGGAAAAGCAGGAGAAA	888
Db	248	AATTTAGTCCAGAAATTTATTTTGTGATGGACCCAGATTTTGGAAAAGCAAGAAGAGA	307
Qy	889	GTGTACAATGTCAATATTCACHTCCACTGGAATGGTACTTTATTTGGAGAAGATCCAGATA	948
Db	308	GTGTACAGATGTCAATACTCACCTCCTTTGGAGTGGTACTTTATTTGGAGAGGATCCGGATA	367
Qy	949	TTTGTCTTAGAGAAATTTGAAGCACAGTGGAGCATTTTCAGCTTTCTGGGAGGGTTTTCAAAA	1008
Db	368	TTTGTCTTAGAGAAATTTAAACACAGTGGAGCGTTCCAGTTGTGTGGGAAGGTTTTCAAAA	427
Qy	1009	GTGGAGAGACAACCTATTCTTCGAGGATTTGTCAATTTGATCCAAAGATGTGTACTCTGTA	1068
Db	428	GTGGAGAAACAATATTCTCTGTAGGATTTGTCAATTTGATCCAAAGCTGTGTGCTGTGTA	487
Qy	1069	TGCACTGCTTCCAGGACAGTGTTCATAAAATCATCGTTTACAAGATGCACTTCTACTG	1128
Db	488	TGGACTGCTTCCAAAGTAGTGTTCATAAAACCATCGTTACAAGATGCACTTCTACTG	547
Qy	1129	GAGGAGGTTCTGTGACTGTGGAGACACAGAGCATGGAAACCTGGCCCTTTTGTGCTAA	1188
Db	548	GAGGGGCTTCTGTGACTGTGGAGACACAGAGCGTGGAAACCTGGCCCTTTTGTGTGG	607
Qy	1189	ATCATGAACCTGGAAGACAGTACTATAAAAGAGAAATTCACGCTGTCCCTTTGAAATGAAG	1248
Db	608	ATCACGAGCCTGGAAGACAGGTACTACAAAAGAGAGCTTACATTGCCCATTCGAATGAAG	667
Qy	1249	AGGTAATTGTCGAAGCCAGGAAATATTTCCCTTCAGTGATAAATATGTGCGTAGAATGA	1308
Db	668	AGGTGATTTGCTCAAGCCAGAGAAATATTCCTTCGGTGATAAAATACATTTGTAGAATGA	727
Qy	1309	CTATATGGGAAGAGGAAAGAAATCGCTCCTGAACCTCCAGATGAAGGAGAAATAAGAAA	1368
Db	728	CTATATGGGAAGAAAGAAAGATGCGCTCCTGAACCTGCAGATGAAGGAGAAATAAGAAC	787
Qy	1369	GATACTATTGTGCTTTTCAATGATGAACACCATTCATATGACCAGCTCATATACAGCC	1428
Db	788	GATACTATTGTGCTTTTCAACGATGAGACCCATTCGTATGATCATGTGATCTACAGTC	847
Qy	1429	TACAAAGAGCTCTTGACTGTGAGCTGCGAGAGGCCAGTTGCAATACACATGCGCATTGACA	1488
Db	848	TGCAGAGAGCTCTAGATTTGGAGCTTGCAGAGGCACAGCTGCACAGCTTGCCATCGACA	907
Qy	1489	AAGAGGTCGTGCGGCTGTTTAAAGCGGAGCTTATGCTGCTTCCAGGAGCAAGCAAG	1548
Db	908	AAGAGGTCGCCGGCTGTCTAAAGCAGGTGTGTATGCCACTTGCCAGGAGCAAGAGG	967
Qy	1549	ATATAAGAGCTCATTCAGAAAATGCTCTCAACATCCACTTCATGTAGAAGTATTACACT	1608
Db	968	ATATAAGAGCTCACTCAGAGACGTCTCTCAGACCCCTCCATGTGGAAGTCTGCACCT	1027
Qy	1609	CAGAGATTATGGCTCATCAGAAAATTTGCTTTGGCTTTGGTTCTTGATGGAACAAATTA	1668
Db	1028	CCGTGGTTATGGCTCACAGAAATTCGCTCTGCGCCTTGCTCTCTGATGAACAAATTA	1087
Qy	1669	TGACTATTCAAGTGACTTTAGGCAGATCTTTTGGCCAAAGCATGCTTTAGAGAGAACCTG	1728
Db	1088	TGAGCTATTCAAGTGACTTTTAGACAGATATTTTGGCAGGCTTGCTCTGTAAGAAACCTG	1147
Qy	1729	ACTCGGAGAAATCCCTGTCTCATAGCAGGTTAATGCTTTGCGATGCAACGCTTTATAAG	1788
Db	1148	GCTCTGAAATCCCTGTCTTATAGCAGACTAATGCTTTGGGATGCAAACTTTTATAAG	1207
Qy	1789	GTGCCCCGAAGATCCCTTCATGAATTCATCTCAGCAGTTTTTTTTTATGGAGATGGAATACA	1848

Db	1208	GTCCCCGTAAAGATCCTTCATGAAATTGATCTTTAGTAGTCTTTTATGGAGATGGAATACA	1261
Qy	1849	AAAAACTCTTTTGGCTATGGAATTTCTGGAAGTATTATAAACAACCTGCAGAAAGAATATATCA	1908
Db	1268	AAAAACTCTTTGGCTATGGAATTTCTGGAAGTATTATAAACAACCTGCAGAAAGAAGTACATCA	1327
Qy	1909	GTGATGATCATGACAGAAAGTATCTCTATAACTGCACCTTTCAGTTCAGATCTTTACTGTTC	1968
Db	1328	CGCAGCACCACGAGAGAAGCATCTCCATAACCGCCCTGTCGGTGCAGATGCTCACCGTCC	1387
Qy	1969	CTACTCTGGCTGCACATCTTATTCGAAGACAGAAATGTTATCTCTGTGCATTTACTTGAACCTC	2028
Db	1388	CGACCTTTGGCCCGGCATCTTATTGAAGACAGAAATGTTATTTCTGTCTCATTTACTTGAACGC	1447
Qy	2029	TGCTAGAAGTTTTACCTGAGTACTTGCACAGGAACAATAAATTCACACTTCCAGGGTTATA	2088
Db	1448	TGCTAGAAGTTTTACCTTGAATACTTTGGACAGGAACAATAAATTCACACTTCCAGGGTTATA	1507
Qy	2089	GCCAGGACAAATTTGGGAAGAGTATATGCAGTAAATATGTGACCTAAAGTATATCCTGATCA	2148
Db	1508	GCCAGGACAAACTGGGAAGAGTCTACGCAGTTATATGTGACCTAAAGTATATCCTGATTA	1567
Qy	2149	GCAAAACCCACAAATGACAGAAAGATTAAGAATGCAGTTCCCTTGAAGGTTTTTCGATCTT	2208
Db	1568	GCAACCCCTGTCAATGACAGAACAGATTAAAGAGCGCAGTTTCCCTGGAAGGTTTTCCGGTCTT	1627
Qy	2209	TTTTTGAAGATTTCTTACCTGTATGCAGGGAATGGAAGAAATCCGAAGACAGGTTTGGCAAC	2268
Db	1628	TTCTGGAAGATTTCTTACCTGTATGCAGGGAATGGAAGAAATCAGAAGACAAGTTTGGACAAC	1687
Qy	2269	ACATTGAAGTGGATCTCTGATTTGGGAGGCTGCCAATGTCTATACAGATGCAATTTGAAGAATA	2328
Db	1688	ACATTGAAGTGGACCCCTGACTGGGAGGCTGCCATCGCTATACAGATGCAACTAAAGAATA	1747
Qy	2329	TTTTTACTCATGTTCCAGAGTGGTGCTCTGTGATGAAGAAGCTTCTTACTTGTGGCTTATA	2388
Db	1748	TTTTTGTCTCATGTTCCAGAGTGGTGCTCTGTGATGAAGATCTCTTACTGGTGGCTTATA	1807
Qy	2389	AAGAATGTCACAAAGCTGTGATGAGGTGCAGTACCAGTTTCATATCTAGTAGCAAGACAG	2448
Db	1808	AAGAATGTCACAAAGCTGTAAAGGTGCAGTACAAATTTCAATGCTCCAGTACCAGACAG	1867
Qy	2449	TAGTACAATCGTGTGGACATAGTTTGGAAACAAAGTCTCTACAGAGTATCTGAGGATCTTGG	2508
Db	1868	TAGTGAATTTGTGCGGTCAATAGTCTGGAACCAAATCCTACAAAGTCTCTGAGGACCTTG	1927
Qy	2509	TAAAGTATACATCTGCCACTCTCTAGGACCCCTTGGCTGTCTTCACTGTACGTTTAAGCAGGC	2568
Db	1928	TAAAGTATACACTGCCACTCTCTAGAACACTTGGCTGTCTTCATGTACGTTTAAAGCAGAC	1987
Qy	2569	TGGGTGCTGTTTCAAGACTGCATCAATTTGTGTCTTTTGAGGACTTTCAAAGTAGAGGTAC	2628
Db	1988	TAGTGTCTATTTCAAGACTGCATGAAATTTGTGCGCTTTTGGACGCTTTCAGTAGAGGTCC	2047
Qy	2629	TAGTGGAAATPCTTTACGTTTGTCTGTTTGGTTGGTCCCAAGTTTGTCTGAGATGTGGC	2688
Db	2048	TGGTGGGATACCCGCTCGCTGCGCTGGTCTGGTGGGCTCAGGTTGTGCTGAGATGTGGC	2107
Qy	2689	GAAGAAATGGACTGCTCTTATTAAGCAGAGTGGTTTTATTACCAAGATGTTTAAGTGCAGAG	2748
Db	2108	GAAGAAACGGGCTCTCACTCATCAGCCAGGTTTTTCTATTATCAAGATGTTTAAATGCAAGG	2167
Qy	2749	AAGAAATGTATGATAAAGATATCATGCTTTCAGATTTGGTCAGATCTTTTAATGATCCCA	2808
Db	2168	AGSAAATGTAGATAAAGATATCATGCTTTCAGATTTGGAGCATCTATTAAATGATCCCA	2227
Qy	2809	ATAAGTTCTTCTTACTTGGTACTTTCAGAGGTATGAACCTTGGCGAGGCTTTTAAACAAGACCA	2868
Db	2228	ACAAGTTCTTCTTACTTGGTACTTTCAGAGATATGAACCTTACTGTATGCTTTTAAACAAGACCA	2287
Qy	2869	TATCTACAAAAGACCAAGATTTGATTTAAACAATAATAATACACTTAATAGAAGAATGCTTC	2928
Db	2288	TATCCACAAAAGACCAAGATTTGATTTAAACAGTATAATACATTAATAGAAGAATGCTTC	2347

Qy	2929	AGTCCCTCATCTATATTGTTGGTGAGCGTTATGTACCTGGAGTGGGAAATGTGACCAG	2988	Qy	4000	AAGAAGAACAGGAGGTGAAATATAGAAATTAATGCCATGGTATTATCGCCCTGTGTCCAGA	4059
Db	2348	AGGTCCTCATCTATATTGTTGGAGAACGTTATGTACCTGGAGTGGGAAATGTACCAGAG	2407	Db	3428	AGAAGAACAGAGAGGTAAAACTAGAAATTAATGCCATGGTATTATCGAGCATGTGTGAGA	3487
Qy	2989	AAGAGGTCACAAATGAGAGAAATCAATTCACTTGTGTTGCATTGAACCCCATGCCACACAGTG	3048	Qy	4060	AATCTACTGCTTTAAACCACAGCAGGGGAAACCCATAGAACTCTCAGGAGAAACCCCTAG	4119
Db	2408	AGGAGGTTAATGAGAGAGATTACTCACTTACTTTGTCATTGAGCCCATGCCACACAGTG	2467	Db	3488	AATCCACCGCCCTTAACCCACAGACAGAGGAAGCCTGTGGACCACTTAAAGGGAACACACTGG	3547
Qy	3049	CCATTGGCCAAAATTTACTTGAGAAATGAAAATATATGAACATGGCTTTAGAGAATGCTATA	3108	Qy	4120	ACCACCTTTTCATGGATCCAGACTTGGCATATGGAACCTTATACAGGAAGCTGTGGTCATG	4179
Db	2468	CCATCGCGAGAACTACTCTGAGAAACGAAATATATGAACATGGCTTTAGAGAATGCTATA	2527	Db	3548	ACCTCTTTTCATGGATCCAGACTTGGCATATGGAACCTTATACAGGAAGCTGTGGTCATG	3607
Qy	3109	ACAAAGTGGCCACATTTAAGAAACCCAGGTGTATCAGGCCATGGAGTTTATGAACATAAG	3168	Qy	4180	TAATGACACGAGTGTGTCGCAAGATTAATTTGAAGCTGTACAGCTGAGCTCTCAGCAGC	4239
Db	2528	ACAAAGTGGCCACATTTAAGAAACCCAGGTGTGTCGGCCATGGAGTTTATGAATTAAG	2587	Db	3608	TAATGATGCAGTGTGCTGGCAGAAGTATTTGAAGCTGTGCACTGAGCTCGCAGCAGC	3667
Qy	3169	ATGAATCACTGAAAGACTTCAATATGTACTTTTATCACTTACTCCTCAAAACCCAGCATGCA	3228	Qy	4240	GCATTGATGTGACCTTTTGGACTTGGAAAGTGAGAAATATCTTTGCCCCTTTTGGCAAT	4299
Db	2588	ATGAATCACTGAAAGACTTCAATATGTACTTTTACCATTATTTCTAAACACAGCATAGCA	2647	Db	3668	GCATTGATGTGACCTTTTGGACTTGGAGAGCGGAGTACCTATGCCCCTCTGCAAGT	3727
Qy	3229	AGGCTGAACATATCAGAGAAAGAGGAGAAACAAAGATGAAGCATTTGCCGC	3288	Qy	4300	CTCTGTGCAATCTGTGATCCCATTAATCTTTGCAACCTCAAAAGATAAACAGTGA	4359
Db	2648	AGGCTGAACATATCAGAGAAAGAGGAGAAACAAAGATGAAGCATTTGCCGC	2707	Db	3728	CTCTGTGCAACACTGTCTCCCATCATCCCTTTGCAAGCGCAGAAAGATCAACAGTGA	3787
Qy	3289	CACCACACCTCCTGAATTTGCCCTGCTTTCAGCAAAAGTGATTAACCTTCTCACTGTG	3348	Qy	4360	ATGCAGATGCTCTTGTCTCAACTTTTGACCCCTGGCAGCGTGGATACACACTGTCTGCCA	4419
Db	2708	CGCCACCTCCTCCAGAGTTCTGCCCCTGCTTTCAGCAAAAGTAGTCAACCTGCTCAGCTGTG	2767	Db	3788	ATGCGGAGGCTCTTGTCTCAACTTTTGACCTTTGGCCGGTGGATACAGACTGTCTTGCCA	3847
Qy	3349	ATATCATGATGTACATTTCTCAGGACCGTATTTTGAGCGGGCAATAGACACAGATTTCTAAT	3408	Qy	4420	GAATATCAGGTTATATAATAAGACATGCTAAAGGAGAAAAACC--CAATTCCTATTCTTCT	4476
Db	2768	ATGTTATGATATACCTCTCAGGACCATCTTTTGAGCGGGCAGTGGACACGGAGTTCTAATC	2827	Db	3848	GAATATCGGGTTATAATATANAACATGCTAAAGAGAGAACCCACAGAGTTCCTGTCTTGT	3907
Qy	3409	TGTGGACCGAAGGATGCTCCAAATGGCTTTTTCATATCTTGGCATTTGGTTTACTAGAG	3468	Qy	4477	TTAATCAAGGAATGGGAGATTTCTACTTTTGGAGTTTCCATTCCTCCTGAGTTTTGGCGTTG	4536
Db	2828	TGTGGACAGAGGATGCTCAGATGGCTTCCATATATTGGCACTGGCTTGTCTGGAG	2887	Db	3908	TTAATCAAGGAATGGGAGATTTCAACTTTTGAGTTTTCATTCCTCCTGAGTTTGGAGTTC	3967
Qy	3469	AGAAGCAACAGCTTCAAAAAGCTCCTGAAGAAAGTAAACATTTTGACTTTTTATCAAGG	3528	Qy	4537	AGTCTTCGATTAATATTCAAATAGCATCAAGGAAATGGTTATTTCTCTTTGGCCACACAA	4596
Db	2888	AGAAGCAGCAGCTTCAGAAAGCTCCTGAAGAGGAAGTGGCTTTTTCACATTTTACCATAAG	2947	Db	3968	AGTCTTCGGTGAATATTCAAATAGTATCAAGGAAATGGTCACTTCTCTGCCACAAACA	4027
Qy	3529	CTTCAAGATTGGGAAGTTCAGCCAT-----GAATATACAAATGCTTTTGGAAAAAC	3579	Qy	4597	TTTATAGAAATTTGAATGAAAGTGGCCACCTGATGAAAGGGATCCTCGAGTCCCCATGCTGA	4656
Db	2948	CTTCAAGATTGGGAAGTTCAGCCATGAATGCTCAGAAATATACAAATGCTTGGAAAGAC	3007	Db	4028	TTTACAGAAATTTGGCCTGAAAGTGGCTCCTGATGAACCTAGACCCACAGAGTGCCCATGATGA	4087
Qy	3580	TCAAAGGAATCCCGAGTTAGAAGGCCAGAGGACATGATAACGTTGGATCTCAGATGT	3639	Qy	4657	CCTGGAGCACCTCGCTTTTCACATATCCAGGCAATTTGAAATCTATTGGGAGATGAAGGAA	4716
Db	3008	TCAAAGGAATCCCGAATTAAGAGGCCAGAGGACATGATACATGGATATCTCCAGATGT	3067	Db	4088	CCTGGAGCAGCTGTGGTTCACCATCCAGCAATCGAAACCTGTTGGGAGATGAAGGAA	4147
Qy	3640	TTGACACAGTGAAGCGATTAAAGAAAAATCTGTTTAATTTGTAGCAACCCACATCAGGAT	3699	Qy	4717	AACCTCTGTTTGGAGCAGCTTCAAAATAGGACGATTAATGTTGAAAGCAFTTAATGCAAT	4776
Db	3068	TTGACACAGTGAAGCGATTAAAGAAAAATCTGTTTAATTTGTAGTGGCAACCACTTCAGGAC	3127	Db	4148	AACCTCTAATTTGGAGCAGCTTCAAAATAGACAGCATAGCGTCTGCAAGCGCTAATGCAAT	4207
Qy	3700	CGGAATCTAATTAAGAAATGATGAGATTACTATGATTAAGAAAAAGCAGAACGAAAAAGAA	3759	Qy	4777	TTGCAAGTTCACAGAGAGATTTACCTGTCTCAGGTCTCTGATACAGAAACATCTGGTTCGTC	4836
Db	3128	TGGAGTGCATTTAAGAGTGAAGAGATTACTCATGATTAAGAAAAAGGCAAGCGGAAGAAA	3187	Db	4208	TTGCAAGTTCACAGAGAGGCTACCTGGCCTCAGGTCTGATACACAAACATCTGGCTCGGC	4267
Qy	3760	AAGCTGAAGCTGTAGGCTACATCGCCAGAGATCATGCTCAGATGTCTGCTTACAGA	3819	Qy	4837	TTCTATCAGTGTGTTCTTCTTAACATAAAAATCAGAAGATACACCATGCTTCTGTCTATAG	4896
Db	3188	AAGCTGAAGCGCGTAGGCTTTCATCGCCAGAGATCATGGCCCAAGATGTGCTCTTACAGA	3247	Db	4268	TCTGTCAAGTATTCTTCTTAACTGCAATCAGAAATAACACAGGCTTCTGTCTGTGG	4327
Qy	3820	AAAACTTCATTGAAACTCATAACTCATGTATGACAATACATCAGAAATGCTGCTGGAAG	3879	Qy	4897	ATCTGTTTCATGTTTGGTGGTGTGTGTAGCATTTCCCATCTTGTATTTGGAGTGAAC	4956
Db	3248	AAAACTTCATTGAAACCCCAAACTCATGTATGATTAATACGTGCAAGTAAACAGGGAAG	3307	Db	4328	ATCTTTCOCATGTTTCTGGTGGCGGAGTCTTAGCGTTTCCCATCTCTTATTTGGATGACA	4387
Qy	3880	AAGATTCCATTATGAGGAAGAGAGCACCAGCAGTCACTGACTTACTAGAAATTTGCTT	3939	Qy	4957	CTGTTGATCTGCAGCCCTTCTTCAAGTGTCTTCTTATTAACCACTTTTATCTTCTTCAT	5016
Db	3308	AAGACTCCATTATGAGGAAGAGAGCACCCTCAGCAGTCACTGAGGSCCTCTAGAAATTTGCTC	3367	Db	4388	CGTGGATCTGCAGCGCTGCCACTTAGTTCTTCATATAACCACTCTATCTCTTCCATC	4447
Qy	3940	TGGTCTCTAAACGGGGTCCATCTGTACTGAAAGGAGGCTGTGACGTGCACTTGTCC	3999	Qy	5017	TGATCACCATGSCACACATGCTTCAGATCTACTTACAGTAGACA-----CAG	5064
Db	3368	TGGGCCCTTAACGGGGCCCGCTGTATTACCGAAAAAGGAGTGTGACGTGCATCTCTCTGCC	3427	Db	4448	TGATCACCATGCGCGACATGCTTCAGATCTCTTACACAGATACAGATCTGTCTCCAG	4507
				Qy	5065	GCCTACCCCTTGCTCAGGTTCAAGAGACAGTGAAGAGGCTCAATTCGCCATCTTCTTCT	5124

Db	4508	GGCGCGCGCTTGCTGAGGGTGAAGAGAGATAGTACAGAGGCTCGCTGTCATCTGCTTTCT	4567
Qy	5125	TTGCAGAAAAATTTCTCAATATACAAGTGGCTTCCATTTGGGTGTGATATTCCCTGGCTGGTATT	5184
Db	4568	TTGTGGAAAGTGTCCGACGACACAGAGCGGCTCACTGGTGGCGGTCTCCCGGCTGGTACC	4627
Qy	5185	TGTGGGTCTCACATGAAGAATGGCATCACCCCTTATCTTCGCTGTGTGCAATTTCTTTTCC	5244
Db	4628	TGTGGCTCTCCCTGAGGAACGGCATACCCCTTTACCTCCGCTGTGTGCACTGCTTTTCC	4687
Qy	5245	ACTATTACTTTGGGTAACTCCGCTGAGGAACATGCATACCAATTTCTGCAGAAGGAGAT	5304
Db	4688	ACTATTACTTTGGAGTAGCTCCGCTGAAGAACTGTTTGGCAATTTCTGCTGAAGAGAAAT	4747
Qy	5305	ACAGTGCACTCTGAGCTATCTATCTTACCTACAAATTTGTTCCCTGCTCTCCAGGAAT	5364
Db	4748	TCAGTGCACTCTGAGCTATCTATCTTACCCACAAATTTGTTCCGCTTTTCCAGGAAT	4807
Qy	5365	ATTGGGATACTGAAGGCCCTTGCTCCAGAGTGGTGTGCAGATCTCTGCCCTTACTAAACT	5424
Db	4808	ATTGGGATACCAATGAAGCCCTTACTACAGAGTGGTGTGGAGATCTCGCCTTACTCAAGT	4867
Qy	5425	GTTTGAAGCAAAAAACACCGTGGTCAGGTACCCCTAGAAAAAGAAATAGTTTGTATAGAGC	5484
Db	4868	CTTTTGAAGCAGAAAAGTGTGTGGTCAGGTACCCCTAGAAAAAGAAATAGTTTGTATAGAGC	4927
Qy	5485	TTCTGTATGACTATAGCTGCCTCTCGAATCAAGCTTCTCATTTTCAGGTGCCACGGTCTG	5544
Db	4928	TTCTGTAGGACTACAGCTGTCTTCTAATCAGGCTTCTCACTTTAGGTGTCCACGGTCTG	4987
Qy	5545	CAGATGATGAGCAAAAGCATCCTGTCTCTGSCCTTTTCTGTGGGGCTATACTATGTCTCTC	5604
Db	4988	CAGATGATGAGCAAAAGCATCCTGTCTCTGTCTTTCTGTGGGGCATCCTGTGTCTCTC	5047
Qy	5605	AGAACATTTGCTGCCAGGAAATTTGTGAACGGGGAAGAGTTTGGAGCTTGCAATTTTTCACG	5664
Db	5048	AGAACATCTGTTGCCAAGAAATATAGTAATGGGGAAGAGTTTGGAGCTGTGGTTTTTCATG	5107
Qy	5665	CACTTCACTGTGGAGCGGAGTCTGCATTTTCTTAAAAATCAGAGATGCCAGTGGTCC	5724
Db	5108	CGCTTCATTTGGTGTGGAGTCTGCATTTTCTTAAAAATCCGAGAAATGACGGGTGGTCC	5167
Qy	5725	TGTTGAAGTGAAGCCAGAGGCTGTGCCTATCCAGCTCCTTACTTTGGATGAATATGGAG	5784
Db	5168	TGTTGAAGGAAGAAGCCAGAGGCTGTGCCTATCCAGCCCTTACTTTGGATGAATATGGAG	5227
Qy	5785	AAACAGACCCTGGCCTGAAGAGGGGCAACCCCTTTCATTTATCTCGTGAAGCGTATCGGA	5844
Db	5228	AAACAGACCAGGGCTAAAGAGAGGAAACCCACTTTCATTTATCTCGGAGCGGTATCGGA	5287
Qy	5845	AGCTCCATTTGGTCTGCAACACACTGCAATTTATAGNAGATTTGTAGAGGCCAAGAGA	5904
Db	5288	AGCTGCATTTGGTCTGGCAACAGCACTGCAATTTAGAAAGATTTGTTCGGAGCCAGGAGA	5347
Qy	5905	CTAATCAGATGTTATTTGGATTCAACTGGCAGTTACTGTGAGCTCCAACCTGCTCCCAAG	5964
Db	5348	CTAATCAGATGCTATTTTGGATTTTAACTGGCAGTATTCTCTGAGCTTCAGTTCTGCCTCAAG	5407
Qy	5965	ACAATCAAAATGACGACAGTAGTAAAGGCTGATTTCAAAATTTATGGAAAACTTTCTGAGG	6024
Db	5408	ACAATCATGATGACATCAATAATAAGACTGATCTAAAAATTTCTAGAGAACTTTCTGAGG	5467
Qy	6025	GCTGGGAAATTTGGAGGCTTTTGTCTCCATGTCCAGGTTCACTTACATCAATATAAT	6084
Db	5468	ACGGGGAAATTTGGAGGCTTTTGTATCCATGTCCAGATTTCTACACACATTTAATAAAT	5527
Qy	6085	ATTCTTTAATGGAGTATTTGCTTTTCAATTAGCAAAACATATGCTTCACAGGAAAAA - AGGAC	6143
Db	5528	ATTCTTTAATGGAAATATTTGCTTTCAATTTCAAAACATTAAGCTTCAAGGGAAAAACAGAC	5587
Qy	6144	ATAGATCAATCTGT-----TTT	6160

Db	5588	ATAGATTAAATGTTTTATGTTTCTAGAACACATAAGAAAAATGCTTGTTCATCCAAAGTGCTAT	5647
Qy	6161	ATGTCGTAGTATTTCCAGGAATTTATTCCTTCATAAATTTGTCCTCATTTTATTTATTT	6220
Db	5648	TTCTGCTAATATTTCCAGAAAACCTCTTCCCTTCATAAATGTCCTAGTTTCATTTTCATAT	5707
Qy	6221	CATCCACTTGGTAGATCAAGTCACGCTCAACACAGTCTGACAGATTTTATGTTGGTTAAC	6280
Db	5708	CACCCACTGTTAATGAGGTCACATTAAGCATTTTGTGGACATTTCTCCATCTGGCTAAC	5767
Qy	6281	TCCTCGCAATTTTGATTTTGGTGT	6307
Db	5768	ATCTCTGCACCTTTGTATTTGGTGT	5794
RESULT 6			
AX714232	AX714232	Sequence 916 from Patent EP1293569.	linear PAT 15-APR-2003
LOCUS	AX714232	3059 bp	DNA
DEFINITION	Sequence 916 from Patent EP1293569.		
ACCESSION	AX714232		
VERSION	AX714232.1	GI:29889184	
KEYWORDS			
SOURCE	homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 Isogai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S., Yamamoto,J.I., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R., Tamechika,I., Seki,N., Yoshikawa,T., Otsuka,M., Nagahari,K. and Masuho,Y.		
TITLE	Full-length cDNAs		
JOURNAL	Patent: Ep 1293569-A 916 19-MAR-2003; Helix Research Institute (JP) ; Research Association for Biotechnology (JP)		
FEATURES	Location/Qualifiers		
source	1..3059		
BASE COUNT	969 a 546 c 694 g 850 t		
ORIGIN			
Query Match	48.4%	Score 3055.8;	DB 6; Length 3059;
Best Local Similarity	99.9%	Pred. No. 0;	
Matches 3057;	Conservative 0;	Mismatches 2;	Indels 0; Gaps 0;
Qy	680	GTCAAGTTTCCTCAAGATGCGGACGAGGAGGCTGGAGGTACTGAGAGGATGGAATCAG	739
Db	1	GTCAAGTTTCCTCAAGATGCGGACGAGGAGGCTGGAGGTACTGAGGAGTGGAAATCAG	60
Qy	740	CGCGGAGTTACCCAGACCCCTCAGCGTCTGGCATCTTGGTGGGATCAGCAAGTTCATTT	799
Db	61	CGCGGAGTTACCCAGACCCCTCAGCGTCTGGCATCTTGGTGGGATCAGCAAGTTCATTT	120
Qy	800	TTTACTGCTTCTTGTCATCTTGGCACAATTTGGTCCAGAAATTTACTTTGCTCAAAAT	859
Db	121	TTTACTGCTTCTTGTCATCTTGGCACAATTTGGTCCAGAAATTTACTTTGCTCAAAAT	180
Qy	860	GGACCCAGACTTGGAAAAGCAGGAGGAAAGTGTACAAATGTCAATATTCACCTCCACTGGA	919
Db	181	GGACCCAGACTTGGAAAAGCAGGAGGAAAGTGTACAAATGTCAATATTCACCTCCACTGGA	240
Qy	920	ATGCTACTTATTTGGAGAACATCCAGATATTTGCTTTAGAGAAATTCAGACACAGTGGAGC	979
Db	241	ATGCTACTTATTTGGAGAACATCCAGATATTTGCTTTAGAGAAATTCAGACACAGTGGAGC	300
Qy	980	ATTTACGCTTTGGGGAGGTTTTCAAAAGTGGAGAGACAAACCTATTCTTGCAGGATTG	1039
Db	301	ATTTACGCTTTGGGGAGGTTTTCAAAAGTGGAGAGACAAACCTATTCTTGCAGGATTG	360
Qy	1040	TGCAATTTGATCCAAACATGTCTACTCTGTATGGACTGCTTCCAGGACAGTGTTCATAAAAA	1099

Db 361 TGCAATTGATCCAACTGTGTACTCTGTATGGACTGCTTCCAGGACAGTGTTCATAAAA 420
QY 1100 TCATCGTTACAGATGCATCTCTACTTGGAGGAGGTTCTGTGACTGTGGAGACACAGA 1159
Db 421 TCATCGTTACAAGATGCATCTCTACTTGGAGGAGGTTCTGTGACTGTGGAGACACAGA 480
QY 1160 GGCATGGAACCTGGCCCTTTTGTGTAAATCATGAACCTGGAGAGCAGTACTATAAA 1219
Db 481 GGCATGGAACCTGGCCCTTTTGTGTAAATCATGAACCTGGAGAGCAGTACTATAAA 540
QY 1220 AGAGAAATTCACGCTGTCGTTGAATGAAGAGGTAATGTGCCAGCCAGGAAATATTTCC 1279
Db 541 AGAGAAATTCACGCTGTCGTTGAATGAAGAGGTAATGTGCCAGCCAGGAAATATTTCC 600
QY 1280 TTCAGTGAATAAATATGCTGTAGAAATGACTATATGGAGAGGAAAGAACTGCCCTCC 1339
Db 601 TTCAGTGAATAAATATGCTGTAGAAATGACTATATGGAGAGGAAAGAACTGCCCTCC 660
QY 1340 TGAACCTCCAGATAAGGGAGAAAATGAAGATACTATTGTGCTTTTCAATGATGAACA 1399
Db 661 TGAACCTCCAGATAAGGGAGAAAATGAAGATACTATTGTGCTTTTCAATGATGAACA 720
QY 1400 CCATTCATATGACCACGTCATATACAGCCCTACAAAGAGCTCTTCACTGTGAGCTCGCAGA 1459
Db 721 CCATTCATATGACCACGTCATATACAGCCCTACAAAGAGCTCTTCACTGTGAGCTCGCAGA 780
QY 1460 GGGCCAGTTGCATACCACTGCCATGTACAAAGAGGTCGTGGGCTGTTAAAGGGGAGC 1519
Db 781 GGGCCAGTTGCATACCACTGCCATGTACAAAGAGGTCGTGGGCTGTTAAAGGGGAGC 840
QY 1520 TTATGCTGCTGCCAGGAACGAAGATATAAAGAGTCATTCAGAAAATGCTCTCA 1579
Db 841 TTATGCTGCTGCCAGGAACGAAGATATAAAGAGTCATTCAGAAAATGCTCTCA 900
QY 1580 ACATCCACTTCATGTAGAAGTATTACACTCAGAGATTATGGCTCATCAGAAAATTTGCTTT 1639
Db 901 ACATCCACTTCATGTAGAAGTATTACACTCAGAGATTATGGCTCATCAGAAAATTTGCTTT 960
QY 1640 GGGCTTTGGTCTCGATGAACAAAATTTAGCTATTCAAGTGACTTTAGGAGATCTT 1699
Db 961 GGGCTTTGGTCTCGATGAACAAAATTTAGCTATTCAAGTGACTTTAGGAGATCTT 1020
QY 1700 TTGCCAAGCATGCCTTAGAGAACAACCTGACTCGGAGATCCCTGCTCATAGCAGGTT 1759
Db 1021 TTGCCAAGCATGCCTTAGAGAACAACCTGACTCGGAGATCCCTGCTCATAGCAGGTT 1080
QY 1760 AATGCTTTGGGATGCAAGCTTTATAAAGTGGCCGTAAAGTCCCTTCATCAATTTGATCTT 1819
Db 1081 AATGCTTTGGGATGCAAGCTTTATAAAGTGGCCGTAAAGTCCCTTCATCAATTTGATCTT 1140
QY 1820 CAGCAGTTTTTTTATGGAGATGGAATACAAAAAACHCTTTTGCTATGGAATTTGTGAAGTA 1879
Db 1141 CAGCAGTTTTTTTATGGAGATGGAATACAAAAAACHCTTTTGCTATGGAATTTGTGAAGTA 1200
QY 1880 TTATAAACAATGCGAGAAGAATATATCAGTGATGATCATGACAGAAGTATCTCTATAAC 1939
Db 1201 TTATAAACAATGCGAGAAGAATATATCAGTGATGATCATGACAGAAGTATCTCTATAAC 1260
QY 1940 TGCACCTTCAGTTCAGATGTTTACTGTTCTTACTCTGGCTCGACATCTTATGAGAGCA 1999
Db 1261 TGCACCTTCAGTTCAGATGTTTACTGTTCTTACTCTGGCTCGACATCTTATGAGAGCA 1320
QY 2000 GAATGTTATCTGTCTACTTACAAACTCTGTCTAGAAGTTTTTACCTGAGTCTGGACAG 2059
Db 1321 GAATGTTATCTGTCTACTTACAAACTCTGTCTAGAAGTTTTTACCTGAGTCTGGACAG 1380
QY 2060 GAACATAAATCAACTTCAGGGTTTATAGCCAGGACAAATTTGGGAGAGTATATGCGAGT 2119
Db 1381 GAACATAAATCAACTTCAGGGTTTATAGCCAGGACAAATTTGGGAGAGTATATGCGAGT 1440
QY 2120 AATATGTGACCTAAAGTATATCTGTATGACGAAACCCCAATATGACAGAGAAATTAAG 2179
Db 1441 AATATGTGACCTAAAGTATATCTGTATGACGAAACCCCAATATGACAGAGAAATTAAG 1500

QY 2180 AATCAGTTCTTGAAGGTTTTTCGATCTTTTTTGAAGATTTCTTACCTGTATGTCAGGAAAT 2239
Db 1501 AATCAGTTCTTGAAGGTTTTTCGATCTTTTTTGAAGATTTCTTACCTGTATGTCAGGAAAT 1560
QY 2240 GGAAGAAATCCGAAGACAGAGTTGGGCAACACATTTGAAGTGGATCTCTGATTTGGAGGCTGC 2299
Db 1561 GGAAGAAATCCGAAGACAGAGTTGGGCAACACATTTGAAGTGGATCTCTGATTTGGAGGCTGC 1620
QY 2300 CATTTGCTATACAGATGCAATTTGAAGAATATTTTACTCATGTTTCCAGAGTGTGTGCTTG 2359
Db 1621 CATTTGCTATACAGATGCAATTTGAAGAATATTTTACTCATGTTTCCAGAGTGTGTGCTTG 1680
QY 2360 TGATGAAGACCTTTACTTGTGGCTTATAAAGAATGTACAAAGCTGTGATGAGTGCAG 2419
Db 1681 TGATGAAGACCTTTACTTGTGGCTTATAAAGAATGTACAAAGCTGTGATGAGTGCAG 1740
QY 2420 TACCAGTTTCATATCTAGTAGCAAGACAGTAGTAGTACAATCGTGTGACATAGTTTGAAC 2479
Db 1741 TACCAGTTTCATATCTAGTAGCAAGACAGTAGTAGTACAATCGTGTGACATAGTTTGAAC 1800
QY 2480 AAAGTCTACAGAGTATCTGAGGATCTTTGAAGCATACATCTGCCACTCTCTAGGACCT 2539
Db 1801 AAAGTCTACAGAGTATCTGAGGATCTTTGAAGCATACATCTGCCACTCTCTAGGACCT 1860
QY 2540 TGCTGGTCTTCATGTACGTTTTAAGCAGGCTGGGTGCTGTTTCAAGACTGCATGATTTGT 2599
Db 1861 TGCTGGTCTTCATGTACGTTTTAAGCAGGCTGGGTGCTGTTTCAAGACTGCATGATTTGT 1920
QY 2600 GTCTTTTGAGGACTTTCAAGTAGAGTACTAGTGAATATCTTTACGTTGCTGCTGTT 2659
Db 1921 GTCTTTTGAGGACTTTCAAGTAGAGTACTAGTGAATATCTTTACGTTGCTGCTGTT 1980
QY 2660 GGTTCGCCAGGTTGCTGTGAGATGTGGCGAAGAAATGACTGTCTCTTATTAGCCAGT 2719
Db 1981 GGTTCGCCAGGTTGCTGTGAGATGTGGCGAAGAAATGACTGTCTCTTATTAGCCAGT 2040
QY 2720 GTTTTATTACCAAGATGTTAAGTGCAGAGAGAAATGTATGATAAGATATCATCATGCT 2779
Db 2041 GTTTTATTACCAAGATGTTAAGTGCAGAGAGAAATGTATGATAAGATATCATCATGCT 2100
QY 2780 TCAGATTTGGTGCATCTTTAATGGATCCCAATAAGTTCTTGTACTGGTACTTTCAGAGGTA 2839
Db 2101 TCAGATTTGGTGCATCTTTAATGGATCCCAATAAGTTCTTGTACTGGTACTTTCAGAGGTA 2160
QY 2840 TGAACCTGCCAGGCTTTTAAACAGACCATATCTACAAAGACCCAGGATTTGATTAACA 2899
Db 2161 TGAACCTGCCAGGCTTTTAAACAGACCATATCTACAAAGACCCAGGATTTGATTAACA 2220
QY 2900 ATATAATACATAAGAAATAGCTTTCAGTCTCTCATATATGTTGGGTGAGCGTTA 2959
Db 2221 ATATAATACATAAGAAATAGCTTTCAGTCTCTCATATATGTTGGGTGAGCGTTA 2280
QY 2960 TGTACCTGGAGTGGAAATGTGACCAAGAGAGGTCAACATGAGAGAAATCAATTCAC 3019
Db 2281 TGTACCTGGAGTGGAAATGTGACCAAGAGAGGTCAACATGAGAGAAATCAATTCAC 2340
QY 3020 GCTTTGCTATCAACCCATGCCACAGTGCATTCGCCAAAATTTTACCTGAGAATGA 3079
Db 2341 GCTTTGCTATGAACCCATGCCACAGTGCATTCGCCAAAATTTTACCTGAGAATGA 2400
QY 3080 TAATGAACTGGCTTAGAGAAATGTATAAACAAAGTGGCCACATTTAAGAAACCCAGG 3139
Db 2401 TAATGAACTGGCTTAGAGAAATGTATAAACAAAGTGGCCACATTTAAGAAACCCAGG 2460
QY 3140 ATCAGGCCATGGAGTTTATGAACATAAAGATGAATCACTGAAAGACTTCAATATGACT 3199
Db 2461 ATCAGGCCATGGAGTTTATGAACATAAAGATGAATCACTGAAAGACTTCAATATGACT 2520
QY 3200 TTATCATTTACTCCAAAACCCAGCATAGCAAGGCTGAACATATGACAGAGAAAGAGAA 3259
Db 2521 TTATCATTTACTCCAAAACCCAGCATAGCAAGGCTGAACATATGACAGAGAAAGAGAA 2580

|||||
841 TTATGCTGCTTGCAGGAAGCAAGAGAGATATAAAGAGTCATTCAGAAAATGTCCTCA 900
QY 1580 ACATCCACTTCATGTAAGAGTATTACACTCAGAGATTATGGCTCATCAGAAAATGCTTT 1639
Db 901 ACATCCACTTCATGTAAGAGTATTACACTCAGAGATTATGGCTCATCAGAAAATGCTTT 960
QY 1640 GCGTCTTGCTTCTGATGACAAAATATAGACTATTCAAGTGACTTTAGGCGAGATCTT 1699
Db 961 GCGTCTTGCTTCTGATGACAAAATATAGACTATTCAAGTGACTTTAGGCGAGATCTT 1020
QY 1700 TTGCCAAGCATGCTTTAGAGAAACCTGACTCGGAGAAATCCCTGCTCTCATAGCAGGTT 1759
Db 1021 TTGCCAAGCATGCTTTAGAGAAACCTGACTCGGAGAAATCCCTGCTCTCATAGCAGGTT 1080
QY 1760 AATGCTTTGGGATGCAAGCTTTATAAGAGTGCCCGTAAGATCCTTCATGAATTGATCTT 1819
Db 1081 AATGCTTTGGGATGCAAGCTTTATAAGAGTGCCCGTAAGATCCTTCATGAATTGATCTT 1140
QY 1820 CAGCAGTCTTTTATGGAGATGGAATACAAAAAATCTTTGGCTATGGAATTTGGAAGTA 1879
Db 1141 CAGCAGTCTTTTATGGAGATGGAATACAAAAAATCTTTGGCTATGGAATTTGGAAGTA 1200
QY 1880 TTATAAACACTGCAGAAAGATATATCAGTGATGATCATGACAGAGATCTCTATAAC 1939
Db 1201 TTATTAAACACTGCAGAAAGATATATCAGTGATGATCATGACAGAGATCTCTATAAC 1260
QY 1940 TGCACCTTTTCAGTTCAGATGTTACTGTTCTCTACTCTGGCTCGACATCTTATTGAAGAGCA 1999
Db 1261 TGCACCTTTTCAGTTCAGATGTTACTGTTCTCTACTCTGGCTCGACATCTTATTGAAGAGCA 1320
QY 2000 GAATGTTATCTGTCATTAAGAACTCTGCTAGAAAGTTTACCTGAGTACTTGGACAG 2059
Db 1321 GAATGTTATCTGTCATTAAGAACTCTGCTAGAAAGTTTACCTGAGTACTTGGACAG 1380
QY 2060 GAACATAAATCAACTCCAGGTTATAGCCAGGACAAATTTGGAGAGTATATGCAGT 2119
Db 1381 GAACATAAATCAACTCCAGGTTATAGCCAGGACAAATTTGGAGAGTATATGCAGT 1440
QY 2120 AATATGTGACCTAAAGTATATCTGATCAGCAAAACCCACAAATATGACAGAAAATTAAG 2179
Db 1441 AATATGTGACCTAAAGTATATCTGATCAGCAAAACCCACAAATATGACAGAAAATTAAG 1500
QY 2180 AATGAGTCTCTGTAAGGTTTTCGATCTTTTGAAGATCTTTACCTGATGACAGGAAT 2239
Db 1501 AATGAGTCTCTGTAAGGTTTTCGATCTTTTGAAGATCTTTACCTGATGACAGGAAT 1560
QY 2240 GGAAGNAATCCGAGACAGTTGGGCAACACATTTGAAGTGGATCTTGGGAGGCTGC 2299
Db 1561 GGAAGNAATCCGAGACAGTTGGGCAACACATTTGAAGTGGATCTTGGGAGGCTGC 1620
QY 2300 CATTTGCTATACAGATCAATTTGAAGAAATATTTTACTCATGTTCCAAAGAGTGGTGTCTTG 2359
Db 1621 CATTTGCTATACAGATCAATTTGAAGAAATATTTTACTCATGTTCCAAAGAGTGGTGTCTTG 1680
QY 2360 TGATGAAGAACTCTTACTTTGGCTTTATAAAGAAATGTCACAAAGCTGTGATGAGGTCAG 2419
Db 1681 TGATGAAGAACTCTTACTTTGGCTTTATAAAGAAATGTCACAAAGCTGTGATGAGGTCAG 1740
QY 2420 TACCAGTTTTCATATCTAGTAGCAAGACAGTAGTACAATCGTGGGACATAGTTTGGAAAC 1800
Db 2480 AAAGTCTTACAGATATCTGAGAGATCTTGAAGCATACATCTGCCACTCTCTAGGACCT 2539
QY 1801 AAAGTCTTACAGATATCTGAGAGATCTTGAAGCATACATCTGCCACTCTCTAGGACCT 1860
Db 2540 TCGTGTCTTTCATGTTACGTTTAAAGCAGGCTGGTGTCTTCAAGACTGCATGAATTTGT 2599
QY 1861 TCGTGTCTTTCATGTTACGTTTAAAGCAGGCTGGTGTCTTCAAGACTGCATGAATTTGT 1920
Db 2600 GTCTTTTGGAGACTTTCAAGTAGAGGTACTAGTGGAAATCTCTTACGTTGTCTGTGTT 2659
|||||

Db 1921 GTCTTTTGAGACCTTTCAAGTAGAGGTACTAGTGAATATCCTTACGTTGTCTGGTGT 1980
QY 2660 GGTTCGCCAGGTTGTGCTGAGATGTGGCGAAGAAATGGACTGTCTCTTATTAGCCAGT 2719
Db 1981 GGTTCGCCAGGTTGTGCTGAGATGTGGCGAAGAAATGGACTGTCTCTTATTAGCCAGT 2040
QY 2720 GTTTTATTACCAAGATGTTAAGTGACAGAGAAGAAATGATGATAAAGATATCATGCT 2779
Db 2041 GTTTTATTACCAAGATGTTAAGTGACAGAGAAGAAATGATGATAAAGATATCATGCT 2100
QY 2780 TCAGATTGGTGCATCTTTAATGGATCCCAATAAGTCTTGTGTTACTGGTACTTCAGAGTA 2839
Db 2101 TCAGATTGGTGCATCTTTAATGGATCCCAATAAGTCTTGTGTTACTGGTACTTCAGAGTA 2160
QY 2840 TGAACCTTCCGAGGCTTTTAAACAAGACCATATCTACAAAGACCAGGATTTGATTAAACA 2899
Db 2161 TGAACCTTCCGAGGCTTTTAAACAAGACCATATCTACAAAGACCAGGATTTGATTAAACA 2220
QY 2900 ATATAATACACTAATAGAAGAAATGCTTCAGGTCCTCATCTATATTGTTGGTGAGGTTA 2959
Db 2221 ATATAATACACTAATAGAAGAAATGCTTCAGGTCCTCATCTATATTGTTGGTGAGGTTA 2280
QY 2960 TGTACCTGGAGTGGGAAATGTACCAAAAGAGAGGTCAAAAGAGAGAAATCATTCACCT 3019
Db 2281 TGTACCTGGAGTGGGAAATGTACCAAAAGAGAGGTCAAAAGAGAGAAATCATTCACCT 2340
QY 3020 GCTTTGCTATTGAACCCATGCCACAGTGCCTTCCAAAATTTACCTGAGAATGAAA 3079
Db 2341 GCTTTGCTATTGAACCCATGCCACAGTGCCTTCCAAAATTTACCTGAGAATGAAA 2400
QY 3080 TAATGAACTGGCTTTAGAGAAATGTCATAAACAAGTGCGCCACATTTAAGAAAACCGAGT 3139
Db 2401 TAATGAACTGGCTTTAGAGAAATGTCATAAACAAGTGCGCCACATTTAAGAAAACCGAGT 2460
QY 3140 ATCAGGCCATGGGTTTATGAACATAAAGATGAATCACTGGAAGACTTCAATATGACTTT 3199
Db 2461 ATCAGGCCATGGGTTTATGAACATAAAGATGAATCACTGGAAGACTTCAATATGACTTT 2520
QY 3200 TTATCATTTACTTCCAAAACCCAGCATAGCAAGGCTGAACATATGCAAGAAAAGAGAAA 3259
Db 2521 TTATCATTTACTTCCAAAACCCAGCATAGCAAGGCTGAACATATGCAAGAAAAGAGAAA 2580
QY 3260 ACAAGAAAACAAAGATGAAGCATTTGCCGCCACCACCTCTGTAATTTGCCCTGCTTT 3319
Db 2581 ACAAGAAAACAAAGATGAAGCATTTGCCGCCACCACCTCTGTAATTTGCCCTGCTTT 2640
QY 3320 CAGCAAGATGATTAACCTTCTCAACTGTGATGATCATGATGATGATGATGATGATGAT 3379
Db 2641 CAGCAAGATGATTAACCTTCTCAACTGTGATGATGATGATGATGATGATGATGATGAT 2700
QY 3380 TGACGGGCAATAGACACAGATTTCTAAGTGTGGACCGAAGGATGCTTCCAAATGGCTTT 3439
Db 2701 TGACGGGCAATAGACACAGATTTCTAAGTGTGGACCGAAGGATGCTTCCAAATGGCTTT 2760
QY 3440 TCATATTTCTGGCATTTGGGTTTACTAGAGAGAAAGCAACAGCTTCAAAAGCTTCCGAAGA 3499
Db 2761 TCATATTTCTGGCATTTGGGTTTACTAGAGAGAGAAAGCAACAGCTTCAAAAGCTTCCGAAGA 2820
QY 3500 AGAAGTAACATTTGACTTTTATCATAGGCTTCAAGATTTGGAAAGTTCAGCCATGAATAT 3559
Db 2821 AGAAGTAACATTTGACTTTTATCATAGGCTTCAAGATTTGGAAAGTTCAGCCATGAATAT 2880
QY 3560 ACAAATGCTTTTGGAAAACCTCAAGGAAATTTCCCGAGTTAGAAGCCAGAGGACATGAT 3619
Db 2881 ACAAATGCTTTTGGAAAACCTCAAGGAAATTTCCCGAGTTAGAAGCCAGAGGACATGAT 2940
QY 3620 AAGTGTGATATCTCAGATGTTTGACACAGTGAAGGATTAAGAGAAAATCTTGTGTTAAT 3679
Db 2941 AAGTGTGATATCTCAGATGTTTGACACAGTGAAGGATTAAGAGAAAATCTTGTGTTAAT 3000
QY 3680 TGTAGCAACCATATCAGGATCGGATCTATTAAAGAAATGATGAGATTTACTCATGATAAAG 3738
Db 3001 TGTAGCAACCATATCAGGATCGGATCTATTAAAGAAATGATGAGATTTACTCATGATAAAG 3059
|||||

QY	2689	GAGAAATGACCTGCTCTCTATTAGCCAGCGTGTATTTATACCAAGATGTTAAAGTGCAGAG	2748
Db	1501	GAGAAATGACCTGCTCTCTATTAGCCAGCGTGTATTTATACCAAGATGTTAAAGTGCAGAG	1560
QY	2749	AAGAAATGATGATAAAGATATCATCATGCTTCAGATTGCTGATCTTTAATGATGCCCA	2808
Db	1561	AAGAAATGATGATAAAGATATCATCATGCTTCAGATTGCTGATCTTTAATGATGCCCA	1620
QY	2809	ATAAGTCTTGTTACTGTTACTTTCAGAGGTATGAACCTTGCAGGCGCTTTTAAACAGACCA	2868
Db	1621	ATAAGTCTTGTTACTGTTACTTTCAGAGGTATGAACCTTGCAGGCGCTTTTAAACAGACCA	1680
QY	2869	TATCTACAAAGACCGAGATTTCATTAACAATAATATACATCAATAGAGAAATGCTTC	2928
Db	1681	TATCTACAAAGACCGAGATTTCATTAACAATAATATACATCAATAGAGAAATGCTTC	1740
QY	2929	AGTCTCTCATCTATATTGCTGGGTGAGGTTATGTACCTGAGTGGGAAATGTGACCAAG	2988
Db	1741	AGTCTCTCATCTATATTGCTGGGTGAGGTTATGTACCTGAGTGGGAAATGTGACCAAG	1800
QY	2989	AAGAGGTACAATGAGAGAAATCATTCACCTTGCTTTTGCAATGAACCCATGCCACACAGTG	3048
Db	1801	AAGAGGTACAATGAGAGAAATCATTCACCTTGCTTTTGCAATGAACCCATGCCACACAGTG	1860
QY	3049	CCATTGCCAAAATTTACCTGAGATGAATAATGAATGAATGCTTGAAGATGTCATAA	3108
Db	1861	CCATTGCCAAAATTTACCTGAGATGAATAATGAATGAATGCTTGAAGATGTCATAA	1920
QY	3109	ACAAAGTGGCCACATTTAGAAACCCAGGTGTATCAGGCCATGGATTTATGAACATAAG	3168
Db	1921	ACAAAGTGGCCACATTTAGAAACCCAGGTGTATCAGGCCATGGATTTATGAACATAAG	1980
QY	3169	ATGAATCACTGAAGACCTCAATATGTACTTTTATCATTTACTCCAAAACCCAGCATAGCA	3228
Db	1981	ATGAATCACTGAAGACCTCAATATGTACTTTTATCATTTACTCCAAAACCCAGCATAGCA	2040
QY	3229	AGCTGAACATATGCAGAGAAAGAGAGAAACAAAGAAATGAAGATGAAGATGCGCG	3288
Db	2041	AGCTGAACATATGCAGAGAAAGAGAGAAACAAAGAAATGAAGATGAAGATGCGCG	2100
QY	3289	CACCACCACTCTGATCTGCTGCTGCTTTTCAGCAAGTGAATTAACCTTCTCAACTGTG	3348
Db	2101	CACCACCACTCTGATCTGCTGCTGCTTTTCAGCAAGTGAATTAACCTTCTCAACTGTG	2160
QY	3349	ATATCATGATGATCTCTCAGGACCGTATTTGAGCGGGCAATAGACACAGATTTCTAACT	3408
Db	2161	ATATCATGATGATCTCTCAGGACCGTATTTGAGCGGGCAATAGACACAGATTTCTAACT	2220
QY	3409	TGTGGACCGAAGGATGCTCCAAATGGCTTTTCATATTTCTGGCATTTGGGTTTACTAGAAG	3468
Db	2221	TGTGGACCGAAGGATGCTCCAAATGGCTTTTCATATTTCTGGCATTTGGGTTTACTAGAAG	2280
QY	3469	AGAAGCAACAGCTTCAAAAAGCTCTCGAAGAAGTAACATTTGACTTTTATCATAAAG	3528
Db	2281	AGAAGCAACAGCTTCAAAAAGCTCTCGAAGAAGTAACATTTGACTTTTATCATAAAG	2340
QY	3529	CTTCAAGATTGGAAGTTCAGCCATGAATATACAAATGCTTTTGGAAAACTCAAGGAA	3588
Db	2341	CTTCAAGATTGGAAGTTCAGCCATGAATATACAAATGCTTTTGGAAAACTCAAGGAA	2400
QY	3589	TTCCCACTTTAGAAGGCCAAGGACATGATGAACGTTGATCTTCAGATGTTTGACACACAG	3648
Db	2401	TTCCCACTTTAGAAGGCCAAGGACATGATGAACGTTGATCTTCAGATGTTTGACACACAG	2460
QY	3649	TGAAGCGATTAGAGAAAAATCTTGTAAATTTAGCAACACACATCAGGATCGGAATCTA	3708
Db	2461	TGAAGCGATTAGAGAAAAATCTTGTAAATTTAGCAACACACATCAGGATCGGAATCTA	2520
QY	3709	TTAAGAAATGATGATGATCTCATGATAAG	3738
Db	2521	TTAAGAAATGATGATGATCTCATGATAAG	2550

RESULT 9	AK027803	2550 bp	mrna	linear	PRI 01-AUG-2002
LOCUS	AK027803	2550	2550	2550	2550
DEFINITION	Homo sapiens CDNA FLJ14897 fis, clone PLACE1004743, weakly similar to PROBABLE N-END-RECOGNIZING PROTEIN.				
ACCESSION	AK027803	GI:14042751			
VERSION	AK027803.1	GI:14042751			
KEYWORDS	oligo capping; fis (full insert sequence).				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Aotsuka,S., Yoshikawa,Y., Matsunawa,H., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuho,Y. and Sasaki,N.				
TITLE	NEDO human cDNA sequencing project				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 2550)				
AUTHORS	Isogai,T. and Otsuki,T.				
TITLE	Direct Submission				
JOURNAL	Submitted (10-MAY-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection: Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.				
FEATURES	Location/Qualifiers				
SOURCE	1..2550				
	/organism="Homo sapiens"				
	/mol_type="mrna"				
	/db_xref="taxon:9606"				
	/clone="PLACE1004743"				
	/tissue_type="placenta"				
	/clone_lib="PLACE1"				
	/note="cloning vector: pME18SFL3"				
	117..>2550				
	/code="unnamed protein product"				
	/codon_start=1				
	/protein_id="BAB55380.1"				
	/db_xref="GI:14042752"				
	/translation="MTIWEKEKLPPELQIRKERNRYCYVLNDEHSHYDHVYSLQR ALDCELAQAQLHTTAIDKEGRVAVKAGAYACQAEKEDIKSHENVSOHPLHVEVLHS EIMARQFALRGLSMNKMIMSYSSDFRQIFCQACLRPEEDSENPCILSRMLMDAKLY KGARKILHELIFSFFMEMEYKLFAMEFVKYKLOKEYISDDHRSISITALSVMQ PTVPTLARHLIEONVISVITETLLEVLPEYLDNRKNFQGYSDQKLGRAVAVICDL KYILISKPTIWRRLRMQFLEGFRSLKILTCMQMEEIRROVQGHIEVDPPWEAIA IQMQLNLLMFQWCACDELLIYAKCHAVKRCSTFSISSRTVYVQSCGHSLET KSYRVEDLVSIHLPLSRLAGLHVRLGASRLHEFVSDFQVEVLVYPLRCL VLVAAVVAEMRRRLSLSIOVFYQDVKCRREMYDKDIIMLQIGASLMDPNKFLILV LQRYELAEAKNTISTKQDLIKOYNTLIEMLQVLIYIGERYVPGVGNVTKVEVTM REIHLICIEPMPHSAIAKNLPENNETGLENVINKVATFKKPGVSGHGYELADES LKDFNMFTYHYSKTOHSAEHMQRKQKQNDKDEALPPPPPEFCAPSKVILNLLCD IMYVILRTVFERAIDTNSLWTEGMLQAFHIALGLLEEKQOLQKAPSEEVTFDFYH KASRLGSSAMNDKIMLEKLGITPQLEGQDKMITWILQMFDTVKRLREKSLIVATTSG SESIKNDRIETHD"				
CDS					
BASE COUNT	830 a	454 c	558 g	708 t	
ORIGIN					
Query Match	40.4%; Score 2546.8; DB 9; Length 2550;				
Best Local Similarity	99.9%; Pred. No. 0;				
Matches 2548; Conservative	0; Mismatches 2; Indels 0; Gaps 0;				
Qy	1189	ATCATGAACCTGGAGAGCAGGTACTATAAAGAGAAATTCACGCTGTCGCTGTAATGAAG	1248		
Db	1	ATCATGAACCTGGAGAGCAGGTACTATAAAGAGAAATTCACGCTGTCGCTGTAATGAAG	60		
Qy	1249	AGGTAATGTCTCAAGCCAGGAAATATTTTCCTTCAGTGATAAAATATGTCGTAGAATGA	1308		

Db	61	 AGGTAATGTGCCAAGCCAGAGAAATATTTCTTTCACTGATAAAATATGCGGTAGAAATGA	120
Qy	1309	CTATATGGGAAGAGGAAAAAAGACTGCCTCCTGAACTCCAGATAAAGGAGAAAAATGAAA	1368
Db	121	CTATATGGGAAGAGGAAAAAAGACTGCCTCCTGAACTCCAGATAAAGGAGAAAAATGAAA	180
Qy	1369	GATACATATGTGTCTTTTCAATGATGAACACACATTCATATGACCACGTCATACAGCC	1428
Db	181	GATACTATTTGTGCTCTTTCAATGATGAACACCATTCATATGACCACGTCATACAGCC	240
Qy	1429	TACAAAGAGCTCTTGACTGTGAGCTGCAGAGGCCAGTTGCATACCACCTGCCATTGACA	1488
Db	241	TACAAAGAGCTCTTGACTGTGAGCTGCAGAGGCCAGTTGCATACCACCTGCCATTGACA	300
Qy	1489	AAGAGGTCCTCGGGCTGTAAAAGCGGGAGCTTATGCTGCTGCCAGGAAGCAAGGAAG	1548
Db	301	AAGAGGTCCTCGGGCTGTAAAAGCGGGAGCTTATGCTGCTGCCAGGAAGCAAGGAAG	360
Qy	1549	ATATAAGAGTCATTCAGAAAAATGTCTCTCAACATCCACTTCATGTAGAGTATTACACT	1608
Db	361	ATATAAGAGTCATTCAGAAAAATGTCTCTCAACATCCACTTCATGTAGAGTATTACACT	420
Qy	1609	CAGAGATTATGGCTCATCAGAAATTTGCTTTGGCTTGGTTCCTGGATGAACAAATTA	1668
Db	421	CAGAGATTATGGCTCATCAGAAATTTGCTTTGGCTTGGTTCCTGGATGAACAAATTA	480
Qy	1669	TGAGCTATTCAAGTGACTTTAGGCAGATCTTTTGGCAAGCATGCCTTAGAGAAGAACCCTG	1728
Db	481	TGAGCTATTCAAGTGACTTTAGGCAGATCTTTTGGCAAGCATGCCTTAGAGAAGAACCCTG	540
Qy	1729	ACTCGGAGAAATCCCTGTCATAGCAGGTTAATGCTTTGGGATGCAAAAGCTTTATAAAG	1788
Db	541	ACTCGGAGAAATCCCTGTCATAGCAGGTTAATGCTTTGGGATGCAAAAGCTTTATAAAG	600
Qy	1789	GTCCCGTAGATCCCTTCATGAATTCATTCAGCAGTTTTTTTTTATGGAGATGGAATACA	1848
Db	601	GTCCCGTAGATCCCTTCATGAATTCATTCAGCAGTTTTTTTTTATGGAGATGGAATACA	660
Qy	1849	AAAAACTCTTTGCTATGGAATTTGTGAAGTATTATAAACAACTGCAGAAAGATATATCA	1908
Db	661	AAAAACTCTTTGCTATGGAATTTGTGAAGTATTATAAACAACTGCAGAAAGATATATCA	720
Qy	1909	GTGATGATCATGACAGAAGTATCTCTATAACTGCACCTTCAGTTCAGATGTTTACTGTTT	1968
Db	721	GTGATGATCATGACAGAAGTATCTCTATAACTGCACCTTCAGTTCAGATGTTTACTGTTT	780
Qy	1969	CTACTCTGGCTGCACATCTTTTGAAGAGCAGAAATGTTATCTCTGTCATTACTGAAACTC	2028
Db	781	CTACTCTGGCTGCACATCTTTTGAAGAGCAGAAATGTTATCTCTGTCATTACTGAAACTC	840
Qy	2029	TGCTAGAAATTTTACCTGAGTACTTGCAGAGGAACAATAAATTCACCTCCAGGGTTATA	2088
Db	841	TGCTAGAAATTTTACCTGAGTACTTGCAGAGGAACAATAAATTCACCTCCAGGGTTATA	900
Qy	2089	GCCAGGACAAATTTGGGAAGAGTATATGCAGTAATATGTGACCTTAAAGTATATCCTGATCA	2148
Db	901	GCCAGGACAAATTTGGGAAGAGTATATGCAGTAATATGTGACCTTAAAGTATATCCTGATCA	960
Qy	2149	GCAACCCCAATATGGACAGAAAGATTAAGAATGCGATTCCTTTGAAGGTTTTTCGATCTT	2208
Db	961	GCAACCCCAATATGGACAGAAAGATTAAGAATGCGATTCCTTTGAAGGTTTTTCGATCTT	1020
Qy	2209	TTTTTGAAGATCTTACCTGTATCGAGGAATGAAGAAATCCGAAGACAGGTTGGGCAAC	2268
Db	1021	TTTTTGAAGATCTTACCTGTATCGAGGAATGAAGAAATCCGAAGACAGGTTGGGCAAC	1080
Qy	2269	ACATTGAAGTGGATCCCTGATTGGAGGCTGCCATTCGTATACAGATGCAATTCGAAGAATA	2328
Db	1081	ACATTGAAGTGGATCCCTGATTGGAGGCTGCCATTCGTATACAGATGCAATTCGAAGAATA	1140
Qy	2329	TTTTTACTCATGTTCCCAAGAGTGGTGTCTGTGATGAAGAACTCTTACTTTGGCTTATA	2388


```
QY 3469 AGAAGCAACAGCTTCAAAAGCTCCTGAAGAAGTAACATTTGACTTTTATCATAGG 3528
Db 2281 AGAAGCAACAGCTTCAAAAGCTCCTGAAGAAGTAACATTTGACTTTTATCATAGG 2340
QY 3529 CTTCAAGATTGGGAAGTTGAGCCATGAATATACAAATGCTTTTGGAAAACTCAAGGAA 3588
Db 2341 CTTCAAGATTGGGAAGTTGAGCCATGAATATACAAATGCTTTTGGAAAACTCAAGGAA 2400
QY 3589 TTCCCAAGTTAGAGGCCAGAGGACATGATAACGTTGGATCTTCAGATGTTTGACACAG 3648
Db 2401 TTCCCAAGTTAGAGGCCAGAGGACATGATAACGTTGGATCTTCAGATGTTTGACACAG 2460
QY 3649 TGAAGCGATTAGAGAAAAATCTTGTAAATTTAGTACCAACCATCAGGATCGGAATCTA 3708
Db 2461 TGAAGCGTTAGAGAAAAATCTTGTAAATTTAGTACCAACCATCAGGATCGGAATCTA 2520
QY 3709 TTAAGAAATGATGAGATTACTTCATGATAAAG 3738
Db 2521 TTAAGAAATGATGAGATTACTTCATGATAAAG 2550

RESULT 10
AY061884
LOCUS
DEFINITION Homo sapiens ubiquitin ligase E3 alpha-II mRNA, complete cds.
ACCESSION AY061884
VERSION AY061884.1 GI:27434479
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 5268)
AUTHORS Han, H.Q. and Kwak, K.
TITLE Novel ubiquitin ligase E3 alpha-II
JOURNAL Unpublished
2 (bases 1 to 5268)
AUTHORS Han, H.Q. and Kwak, K.
TITLE Direct Submission
JOURNAL Submitted (05-NOV-2001) Research, Amgen Inc., One Amgen Center
Drive, Thousand Oaks, CA 91320, USA
FEATURES
source
1..5268
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1..5268
/codon_start=1
/product="ubiquitin ligase E3 alpha-II"
/protein_id="AAL32101.1"
/db_xref="GI:27434480"
/translation="MASELEPEVQALDRSLLECSAEIAGKWLQATDLTREVYQHLAH
YVPKLYCRPNFPFQKDLAHLGLPMEWYLCGEDPAFPFKLPKQANKPSHLGCRV
FYQGFYSCRCADVPDCLVCEMFLGSIHRDHYRMTTSGGGFCDCGDTAEWEG
PKCQHELNTSEIEEDPLVHLSDEVIARTYNIFAIRYAVAILTWKESLPAHL
ENVEKSDTYCMCHNDEVHTYEQIVYTLQKAVNCTOKEAIGFATTVDRGRSVRYGD
FOYCEQAKSVIVRNTSRQPLKQVWHSIVAHQNFGLKLLSLWGLIYSGLRRI
LCQVGLQPGDGENSLDRMLSDSLKMWKARSVTHQLFMSLLMDLIKKLFAVRF
AKNYQLOQRFEDDHERAVSVTALSQVFPTAPTLARMLITEENLMDLIITKTFDHLR
HRDAQGRFERYTLQAFKFRVQSLILDLYVLISKPTESWDELQKFLFEGDAFL
ELKCMQMDHTITROVGOHIEPEWEAAFTLQMKLTHVISMMODWCASDEKVLIEAY
KCLAVLMQCHGYTDGEQPIITLSICGHSVETIRYCVSOEKVSIHLVPSRLLAGHLV
LSKSEVAKFPPELLPLSELSPMLIEHLPLRLCLVLAQVHAGMWRNGFSLVNOIYYH
NVYKREMFDKVMQTCVSMMDPNHLMISRFELYOIFSTDYGRKFSSEITHK
DYVOQNNTLIEMLYLIIMLVGERSPGVQVNDATDEIKREITHLSIKPMHSELVK
SLPEDENKTEGSEVIEVAHFPGKGLTGKRVYELKPECAKENFLYFHFSAEQSKA
EBAORLKLQNRDENTALPPVLPKPPCLFASLVNIIQSDVLMICMTILQWAVEHNGY
ANSENLQVRLHLIGMALQEBQHLNENVEEHVVFYFTQKISKPEAKNPSILAM
LRTLNQAPLYLEHKDMIRWLKTFNAVKMKRESSPTSPVAETEGTLMESSRDKRAE
RRKAEIRAEIRLQMAQSEMRHFIDENKELFOOTLEADSTAVLDHSPVSDMT
LTALGPAQTVPEQRQFVTCILCQEBQEVKVESRAMVLAAFVQVRSVLSKNRSKFIQD
PEKYDPLFPHPLDSCGTHTSSCGHIMHAWCQRYFDSVQAEQRRQRRLRLHTSYDVE
```

```
NGEFLCPLECLSNVTLPLLPRIINRNLFNSDOPNLITOWIRTIISOIKALOFRLK
EESTPNNASTKNSENVDELQLPCEGRPDPRPKIPYSESIKEMLTTFGTATYKVLKWH
FNEEDPRVPIMCNSCATYTIQSIERILSDKPLPCLRDDCLDSLTRFAAAHWT
VASVSVQGHFCKILFASLVNDSHEELPCILDIMFHLVLGLVLAFFAQCODFSGIS
LGTGDLHIHLVMTMAHIIQILITSCTEENGQDENPCEESAVIALYKTLROYTGSAL
LKEIPSGHILWRSVRGIMPFLKCSALFHYLVNVPSPDIOVPGTSHPEHLCSYLSL
PNNLICILPOENSEIIMNSLIESWRNSEVKRYLEGPERDAIRYPRESKNLNLNLPEDYSSL
INQASNSCPKSGGDKSRAPILCLVCGSILCSQSYCCQTELEGEDEVGACATATYSCGS
GWLFLVRRECQVILFAGKTKGFCFYPPYLDYGETDQGLRGNPLHLUCKERPKFIQK
LMHQSVTTEEIGHAEANQTLVGDWQHL"
BASE COUNT 1618 a 1015 c 1193 g 1442 t
ORIGIN
Query Match 21.9%; Score 1378.6; DB 9; Length 5268;
Best Local Similarity 56.4%; Pred. No. 0;
Matches 2945; Conservative 0; Mismatches 2169; Indels 105; Gaps 16;
QY 780 TGGGATCAGCAAGTTGATTTTATATCTCTTCTTGCATCATTTTGCACAAATTTGGTGCA 839
Db 82 TGGCTGCAAGCACTGACCTCAGTACAGAAAGTGTACAGCAATTTAGCCCACTATGTACCC 141
QY 840 GAAATTTACTTTGCTGAAATGGACCACAGCTTGGAAAGACGAGGAGGAAAGTGTACAAATG 899
Db 142 AAATCTACTCCAGGGGTCCCAACCCCTTTTCCACAGAAAGAGACATGCTGCCACAGCAT 201
QY 900 TCAATATTACTTCCACTGGAATGGTACTTATTTTGGAGAAGATGCCAGATATTTGCTTAGAG 959
Db 202 GTTTTGTGGGACCAATGGAATGTACCTTTGTGTGGAAGATCCTGCTGCTTTGGAATTTCCA 261
QY 960 AAATTTGAAGCAC---AGTGGAGCATTTTCAGCTTTGCGGAGGGTTTTCAAAAGTGGAGAG 1016
Db 262 AAATTTGAGCAAGCAAAACAAACCTTCTCATCTTTGTGTCGTGTTTAAAGATAGAGAG 321
QY 1017 ACAACCTATTCTTCAGGGATTGTGCAATTTGATCAACATGTGCTACTCTGTATGACTGCTC 1076
Db 322 CTRACATATTTCCAGAGACTGTGCAGTTGATCCCACTTGTTGTTTGTGCATGGAGTGC 381
QY 1077 TTCAGGACAGTGTTCATAAAATCATCGTTTACAAAGATGCATCTTACTCTGGAGAGGG 1136
Db 382 TTTTGGGAGTATTACAGAGATCATCGATAGATGACAACATCAGGAGGTGGAGGT 441
QY 1137 TTCTGTGACTGTGGAGACACAGAGCATGGAAACCTGGCCCTTTTGTGTAATCATGAA 1196
Db 442 TTCTGTGACTGTGTGACTTGAAGCCTTGAAGAGGGTCTCTTACTGTCAAAAGATGAA 501
QY 1197 CCTGGAAGACAGTACT--ATAAAGAGAAATTCACGCTGTCCCTTGAATCAAGAGGTA 1253
Db 502 CTTAACACCTCTGAAATTTGAGGAAGAGAGGATCCTCTTGTTCATTTATCAGAAGATGG 561
QY 1254 ATTGTCACAGCCAGGAAAAATATTTCCCTTCAGTGATAAAATATGCTGTAAGAAATGACTATA 1313
Db 562 ATAGCAAGAACTTATAACATTTTTCCTATTACGTTTCGGTATGCAAGTAGAATATTAACC 621
QY 1314 TGGGAAGAGGAAAAAGAACTGCCCTCTGAACTCCAGATGAAGGAGAGAAAAATGAAAGATAC 1373
Db 622 TGGAAAAAGAAAAAGTGAATTTGCCAGCAGATTTAGAGATGGTAGAGAAGAGTACACACTAC 681
QY 1374 TATTGTCTCTTTTCAATGATGACACCAATTCATATGACCACGCTCATATACAGCCTACAA 1433
Db 682 TATTGTGCTGTTTAAATGATGAGGTTTACACACCTTATGAACCAAGTATTATTACTCTTCAG 741
QY 1434 AGAGCTCTTACTGCTGAGCTCGCAGGCCGCCAGTTGTCATACACTGCCATTCGCAAAAGAG 1493
Db 742 AAAGCTCTTACTGTACAAAAAGAGAGCTATTGGTTTTCAGTACATAGTAGATCGAGAT 801
QY 1494 GGTGCTCGGGTGTAAAGCGGGAGCTTATGCTGCTTCCGAGGAGGAAAAAGAAAGATATA 1553
Db 802 GGGCGTAGGCTGTGTCGATATGGAGATTTTTCAGTATTGTGAGCAAGCAAAATCAGTAAT 861
QY 1554 AAGAGTCATTCAGAAAAATGCTCTCAACATCCACTTCATGTAGAAAGTATTACATCAGAG 1613
Db 862 GTGAGAAATACCAGTAGACAGAC---AAAGCCACTCAAAGTTCAAAGTTATGCTATCGTCT 918
```


Qy 1614 ATTATGGCTCATCAGAAATTTGCTTTGCTTCTGTTCTTGGTAAACAAATTTATGAGC 1673
Dy 919 ATTGTCGCACATCAGAAATTTGGTTTGAACCTTTTGTCTGGCTGGGAGTATTTATGGA 978
Qy 1674 TATTCAAGTGAATTTAGGAGATCTTTTGGCAAGCATGCCCTTAGAGAAAGACCTGACTCG 1733
Dy 979 TATTTCAGATGCCCTTCGCCGATTTTATGTCAAAGTTGGTTTACAAAGGAGGCGGAGATGGT 1038
Qy 1734 GAGAAATCCCTGCTCATCAAGCAGGTAAATGCTTTGGGATGCCAAAGCTTTATAAAGGTGCC 1793
Dy 1039 GAAACACTCTTCTCTAGTGGACAGACTGATGCTTGTAGTATCCAAATTTATGGAAGAGTGTCT 1098
Qy 1794 CGTAAGATCCCTTCATGAATGATCTTCAGCAGTTTATTTATGAGAGTGAATACAAATAA 1853
Dy 1099 AGGAGTGTATCATCAGTTGTTCTCAGCAGCTGCTTTATGGATTTGAAATACAAAGAA 1158
Qy 1854 CTCCTTTGCTATGGAATTTGTGAAGTATATTAACAACCTGCAGAAAGAAATATATCAGTGTAT 1913
Dy 1159 CTATTTGCTGTTCCGATTTTGGCAAAATTTACCAGCAGTTGCAGAGATTTTATGGAGGAT 1218
Qy 1914 GATCATCAGCAGAAGTATCTCTAATCTGACCTTTTCAGTTTCAGATGTTTACTGTTCCCTACT 1973
Dy 1219 GATCAGCAGGAGCAGTGTGCTGACTGCTCTATCTGTCCAGTTCTTCACCGCACCTACT 1278
Qy 1974 CTGGCTCGACATCTTATTGAGACAGAAATGTTATCTCTGCTCAATTAAGTAACTCTGCTA 2033
Dy 1279 CTGGCTCGAATGCTCATCAGAGAAAGAACTGTATGAGCAATATCATTAAGACATTTTATG 1338
Qy 2034 GAAGTTTTT-----ACCTGAGTACTTGGCAGAGGAACAATAAATTCACCTCCAGGGTTAT 2087
Dy 1339 GATCATTTGAGACATCGAGATGCCAGGCGAGATTTTCAGTTTGAACGATACACTGCTTTA 1398
Qy 2088 AGCCAGGACAAATTTGGGAAGAGTATATGCAATATATGTGACCTTAAAGTATATCTGATC 2147
Dy 1399 CAAGCCTTCAAAATTTAGGAGAGTACAGAGCCCTTATTTTAGATCTCAAGTATGTTAATT 1458
Qy 2148 AGCAAAACCCACAATATGGCAGAGAAGATTAAGATGCACTTCCCTTGAAGGTTTTCGATCT 2207
Dy 1459 AGCAAAACCAATGAATGGTTCAGATGAGCTGAGGAGAGAAATTTCCCTAGAGGGTTTTCATGCC 1518
Qy 2208 TTTTGTGAAGATCTTACTGCTATGTCAGGGAATGGAAGAAATCCGAAGACAGGTTGGGCAA 2267
Dy 1519 TTTTGTGGAATTTACAAATGATATGTCAGGGAATGGATCCAAATACAGCTCAAGTAGGACAA 1578
Qy 2268 CACATTTGAAGTGGATCTCTGATTTGGAGGCTGCCATTTGCTATACAGATCAAAATGGAAGAAT 2327
Dy 1579 CATATTTGAATTTGAACACAGAGTGGAGGAGCCCTTCACACTACAAATGAATTAACACAT 1638
Qy 2328 ATTTTACTCATGTTCCAAAGAGTGTGCTGCTGTTGATGAAGAACTCTTACTTTGCTGCTTAT 2387
Dy 1639 GTCATTTTCAATGATGTCAGGAGTGTGCTGCTTCAGATGAAAAAGTGTAAATCGAAGCTTAC 1698
Qy 2388 AAGCAATGTCAAAAGCTGTGATCAGGTGTCAGTACCAGTTTCATATCTAGTAGCAAGACA 2447
Dy 1699 AAGAAATGTCGCTGTACTGATCGAGTGTGATGGTGTATATCTGATGGTGAACAGGCA 1758
Qy 2448 GT----AGTACAATCGTGTGGACATAGTTTGGAAACAAAGTCCCTACAGAGTATCTGAGGAT 2504
Dy 1759 ATCACACTAAGCAATTTGGACATCTCAGTGGAACTATCAGATACTGTGTTTCCCAAGAA 1818
Qy 2505 CTTGTAAGCATPACATCTGCCACTCTCTAGGACCCTTGGTGTCTTCATGTAGCTTTAAGC 2564
Dy 1819 AAAGTTAGCATTCACCTCCCAAGTTTCTCGCTTACTTGCAGGTTTACATGTATTTAAGC 1878
Qy 2565 AGGCTGGTGTCTTCAAGACTCCATGAATTTGCTGTGTTGGTCCAGGTTGTTGCTCAGATG 2624
Dy 1879 AAAAGTGAAGTGGCATATAAATTTCCAGAGCTCTACCTCTAAGTGAACCTTAGGCCACCC 1938
Qy 2625 GTACTAGTGAATATCTTTACGTTGCTGCTGTTGGTGGTCCAGGTTGTTGCTCAGATG 2684
Dy 1939 ATGTTGATAGAACCCCTCTAGATGCTGTTCTGTGTCGCCAAGTACATGCGCGGAATG 1998
Qy 2685 TGGCGAAGAAATGGACTGCTCTTATTAGCCAGGTTTATTACCAAGATGTTAAGTGC 2744

Dy 1999 TGGAGAAAGAAATGGGTTTCTCTCTAGTAACACAGATTTATTACTACCATTAATGTGAATGC 2058
Qy 2745 AGAGAAGAAATGATGATAAAGATATCATCATCTCTAGATTTGGTGCATCTTTTAATGGAT 2804
Dy 2059 AGACGTGAGATGTTTGGACAAGGATGTAGTAATGCTTCAGACAGAGTGTCTCCATGATGGAT 2118
Qy 2805 CCCAATAAGTCTTCTGTTACTGTTACTTTCAGAGGTATGAACCTTCCCGAGGCTTTTAAACAAG 2864
Dy 2119 CCAAAATCATTTCTCTGATGATCATGCTCAGCCGCTTTGAACTTTATCAGATTTTCACTACT 2178
Qy 2865 ACCATATCTACAAAA-----GACCAGGATTTGATTAACAATAAT 2903
Dy 2179 CCAGACTATGGAAGAAAGATTTAGTTCTGAGATTAACCATATGAGATGTTGTTCCAGCAAG 2238
Qy 2904 AATACACTAATAAGAAAGAAATGCTTCAGGTCTCTCATCTATATTGTTGGGTGAGCGCTTATGTA 2963
Dy 2239 AATACTCTAATAAGAAAGAAATGCTATACCTCATTAATAATGCTTGTGGAGAGATTTAGT 2298
Qy 2964 CTGGAGTGGGAAATGTGACCAAGAAAGAGGTACAAATCAGAGAAATCAATTCACCTGCTT 3023
Dy 2299 CTGGAGTGGGAGAGTAAATGCTACAGATGAATCAAGCAGAGATTTATCCATCAGTTG 2358
Qy 3024 TGCAATTGAACCCATGCCACACAGTCCATTGGCCAAAAATTTACCTCAGAAATGAAAAATAAT 3083
Dy 2359 AGTATCAAGCCCTATGGCTCATAGTAATTTGGTAAAGTCTTTACCTGAAGATGAGAAACAAG 2418
Qy 3084 GAACTGCTTAGAGAAATGCTATAAACAAGTGGCCACATTTTAAAGAAACCAAGTGTATCA 3143
Dy 2419 GAGACTGCGATGGAGAGTGAATCGAAGCAGTTGCCCATTTCAAGAAACCTGGATTAACA 2478
Qy 3144 GGCCATGGAGTTTATGAACATAAAGATGAATCACTGAAAGCTTCAATATGACTTTTAT 3203
Dy 2479 GGACGAGCATGATGAATGAAACAGAAATGTGCCAAAGAGTTCAACTGTTATTTCTAT 2538
Qy 3204 CATTAATCCAAACCCAGCATAGCAGGCTGAACATATCAGAGAGAAAGGAGAGAAACAA 3263
Dy 2539 CACTTTTCAAGGCGAAGACAGTCCAAAGGAGAAAGCGCAACGGAATTTGAAAGACAA 2598
Qy 3264 GAAAAACAAGATGAAGCATTTGCCGCCACCCACCTCTCTGAATTTTCGCCCTGCTTCAGC 3323
Dy 2599 AATAGAGAGATACAGACTCCACCTCCGCTGTTGCCCTCCATTTGCCCTCTGTTGCA 2658
Qy 3324 AAAGTGATTAACCTTTCTCAACTGTGATATCATGATGTATCATTTCTCAGGACCGTATTTAG 3383
Dy 2659 AGCCTGGTTAACAATTTTGCAGTCAGATGTCATGTTGTGTCATCATGGAAACAATTTCTGCAA 2718
Qy 3384 CGGCAANTAGACACAGATCTTAACCTTGTGGACCGAAGGAGTGTCCAATGGCTTTTCAT 3443
Dy 2719 TGGGCTGTGGAAACATAATGGATATGCTGTGTCAGAGTCCATGCTGCAAAAGGTTGTACAT 2778
Qy 3444 ATTTCTGGCATTTGGGTTTACTAGAAGAGAAAGCAACAGCTTCAAAAAGCTCCTGAAG---AA 3500
Dy 2779 TTAATTTGCGATGCGCACTACAGAGAGAAACAAACATTTAGAGAAATGTCAGGAGAGCAT 2838
Qy 3501 GAAGTAAACATTTGACTTTTATCATAAAGGCTTCAAGATTTGGGAAGTTTCAGCCATGAAT--- 3557
Dy 2839 GTAGTAACATTTTACCTTCACTCAGAAGATATCAAAACCTTGGTGAAGCGCAAAAATTTCT 2898
Qy 3558 -----ATACAATGCTTTTGGAAAAACTCAAGGAATTTCCCAAGTTTAGAGGCCAGAG 3611
Dy 2899 CCTAGCATACTAGCTATGCTGGAAACACTACAAAAATGCTCCCTACCTAGAAAGTCCACAA 2958
Qy 3612 GACATGATAACCTGGATCTTACAGATGTTTGACACAGTGAAGCGGATTAAGAGAAATCT 3671
Dy 2959 GACATGATTCGGTGGATATTTGAAGACTTTTAAATGCTGTTTAAAGATGAGGAGAGTTCA 3018
Qy 3672 TGTTTAAATTTAGCAACACATCAGATCGGAATCTTATTAAAGATCATCAGATTTACTCAT 3731
Dy 3019 CCTACCAGTC-----CGTGGCAGAGACAGAGGAAACCATTAATGGAAGAGAGTTCAAGG 3072
Qy 3732 GATAAAGAAAAAGCAGAAACGAAAAAGCTGAAGCTGTAGGCTACATGCTCCCAAG 3791
Dy 3732 GATAAAGAAAAAGCAGAAACGAAAAAGCTGAAGCTGTAGGCTACATGCTCCCAAG 3791

Db	3073	GACAAAGACAAAGCTGAGAGGAGAGAAAGACAGAGATTGCCAGACTGCCAGAGAAAAG	3132
Qy	3792	ATCATGGCTCAGATGCTCGCTTACAGAAAAAAGCTTCATTGAAAGCTCATAAAGCTCATGAT	3851
Db	3133	ATCATGGCTCAGATGCTGAAATGCGAGCGGCATTTTATTGATGAAACAAAGAACTCTTT	3192
Qy	3852	GACATACATCAGAAATGCTGGGAAGAAGATTCATATGAGGAAGAGACACCCCA	3911
Db	3193	CAGCAGACATTTAGAACTGGATGCTCAAGCTCTGCTGTTCTTGATCA-----TAGCCCT	3246
Qy	3912	GCAGTCAGTACTACTAGAAATGCTTTGGGTCCCTAAAGCGGGTCCATCTGTTACTGAA	3971
Db	3247	GTGGCTTCAGATATGACATTACAGCACTGGGCCCGCCACAACTCAGGTTCTCGAACAA	3306
Qy	3972	AAGGAGGTGCTGACGTGCATCTCTTGGCCAGAAAGACAGAGGTGAAATAGAAAAATAT	4031
Db	3307	AGACAATTCGTTACATGATATTGTGTCAAGAGGAGCAAGAACTTTAAAGTGGAAAGCAGG	3366
Qy	4032	GCCATGGTATTATCGGCCCTGTGCCAGAAATCTACTGCCTTAACCCAGACAGGGGAAA	4091
Db	3367	GCAATGGTCTTGGCAGCATTTGTTCAGAGATCAACTGTATTATCAAAAAACAGAAGTAA	3426
Qy	4092	CCCATAGAACTCTCAGGAGAAGCCCTAGACCCCACTTTTCATGATCCAGACTTGGCATAT	4151
Db	3427	TTTATTCAAGATCCAGAAAAA---TATGATCCATTAATTCATGCACCCCTGATCTGTGT	3483
Qy	4152	GGAACTTATACAGGAAGCTGTGTCATGTAATGCACGAGTGTGTCGGCAGAAATATTT	4211
Db	3484	GGAACACACACTAGTAGCTGTGGGCACATTAATGCATGCCATTTGTTGGCAAGGATTTT	3543
Qy	4212	GAAGCTGTACAGCTCTCAGCAGCGCATTCATGTTGACCTTT-----T	4259
Db	3544	GATTCGGTTCAAGCTAAAGAACAGCGAAGGCAACAGAGATTACGCTTACATACAGCATAT	3603
Qy	4260	GACTTGGAAAGTGGAGAATATCTTCCCTCTTTGCAATCTCTGTGCAATCTGTCATCTGATC	4319
Db	3604	GATGTAGAAAAGCGGAATTCCTTTGCCCTTTTGTGAATGCTGTGATTAATCTGTTATT	3663
Qy	4320	CCCATATTTCCTTTGCAACTCAAAAGATAAAGACAGTGAAGATGCAGATGCTTTGCTCAA	4379
Db	3664	CC-----TCTGCTGCTCTCCAGAAATATTTTAAACACAGGTTAAATTTTTCAGAC	3717
Qy	4380	CTTTTGACCTCGCAGGGTGGATACAGATGTTCTGCGCCAGAAATATCAGGTTATATATA	4439
Db	3718	CAACCAAAATCTGACTCAGTGGATTTAGAACAAATATCTCAGCAATAAAGACATTTACAGTTT	3777
Qy	4440	AGACATGCTAAAGGAGAAAACCAATTCCTATTCTTTTAAATCAAGGAATGGGAGATTCT	4499
Db	3778	CTTAGAAGAAGAAGTACTCTTAATAATGCCCTACAAAGAAATTCAGAAATATGGGAT	3837
Qy	4500	ACTTTGGAGTTCCATTCCTGAGTTTTGGCGTTGAGTCTTCGATTAAATATTCAAAT	4559
Db	3838	GAATTACAGCTCCCTGAAGGTTACAGCCCTGATTTTCGTCCTTAAGATCCCTTATCTGAG	3897
Qy	4560	AGCATCAAGAAATGGTTATCTCTTTGGCCAAACAAATTTATAGAAATGGATTGAAAGTG	4619
Db	3898	AGCATAAAAGAAATGCTAACAGACATTTGGAACGTCTACCTACAAAGTGGGACTAAAAGTT	3957
Qy	4620	CCACCTGATGAAGGGATCCCTCGAGTCCCATCTGACCTGGAGCACTCGCGCTTCACT	4679
Db	3958	CATCCCAATTAAGAGGAGCTCTCGTGTCCCTAATAATGTGTGGGTAGCTCGCGGTACAC	4017
Qy	4680	ATCCAGGCAATTTGAAATCTATTGGGAGATGAAGGAAAACCTCTGTTTGGAGCACTTCAA	4739
Db	4018	ATCCAAAGCATAGAAGAATTTTGTAGTGATGAAGATAAACCATTGTTTGGTCTCTTACCT	4077
Qy	4740	AATAGGCAGCATATGGTCTGAAGCATTAATGTCAGTTTTGCAGTTTGCACAGAGGATTACC	4799
Db	4078	TCCAGACTGGATGACTGTCTTAGTTCATTGACGAGATTTTGGCCGACACACACTGGACATG	4137
Qy	4800	TGCTCTCAGGTCCTGATACAGAAACATCTGGTTCGCTCTTCTATCATAGTGTCTTCTCTAAC	4859
Db	4138	GCATCAGTTTTCAGTGTGCAAGGACATTTTGTGTAACATTTTGTGCATCACTGTTGGCTAAT	4197
Qy	4860	ATAAATCAGAAGATACACATGCTTCTGTCTATATAGATCTGTTTCATGTTTGTGGGT	4919
Db	4198	GACAGCCATGAGGAAGTCCATGATATATAGATATTTGACATGTTTTCATTTATTGTTGGGC	4257
Qy	4920	GCTGTGTTAGCATCCCATCTGTTATTTGGGATGACCCCTGTTGATCTCGAGCTTCTTCA	4979
Db	4258	TTGGTGTGTCATTTCTGCGTTGAGTG-----TCAGATTTTTCAGGG	4302
Qy	4980	GTTAGTTCTTCTATAACCACTTTATCTTCTTCCATTTTGATCACCATGGCACATGCTT	5039
Db	4303	ATCAGCCCTGGCACTGGAGACCTTCACATTTTCCATCTGGTTACTATATGSCACACATATA	4362
Qy	5040	CAGATCTACTTACAGTAGACACAGGCC---TACCCCTTGTCTCAGGTTCAAGAAGACAGT	5096
Db	4363	CAGATCTTACTTACTCTCATGTACAGAGAGAAATGGCATGGATCAAGAAAAATCCCTTGT	4422
Qy	5097	GAAGAGCTCATTCCTCGCATCTTCTTCTTCCAGAAATTTCTCAATATACAGTGGCTCC	5156
Db	4423	GAAGAAGTACAGAGTTCTTGTGTTATAAACACTTCACCAAGTATACGGGAAGTGCC	4482
Qy	5157	ATTGGGTGTATATTCCTGCTGGTGTATTTGTGGGTCTCACCTGAAGAAATGGCATCACCCCT	5216
Db	4483	TTGAAGAAGTACCATCCGCTGCGATCTGTGGAGGAGTGTACAGAGCTGGAATCATGCT	4542
Qy	5217	TATCTTCGCTGTGCTGATTTTTCACACTTATTTACTTTGGGTAACCTCCGCTCAGGAA	5276
Db	4543	TTCTGAAAGTGTCTGCTTTATTTTTCATTAATAATGAGTTCCTTCCCAACCCGAC	4602
Qy	5277	CTGCATACCATAATCTGCAGAGGAGATGACAGTGCACCTCTGTAGCTATCTATCTTACCT	5336
Db	4603	ATTCA---AGTTCTCGAACAAGCCATTTTGAACATTTATAGTATCTTTCCCTACCA	4659
Qy	5337	ACAAATTTGTTCTGCTCTTCCAGGAATATTTGGGATACTGTAAAGCCCTTGTCTCAGAGG	5396
Db	4660	AACAACCTCATTTGCCCTTTTCAAGAAATAGTGAAGATAAGAAATTCATGATGAAAGT	4719
Qy	5397	TGCTGTGCAGATCCGCTTACTAACTGTTTGAAGCAAAAAACACCGTGTGCTCAGGTAC	5456
Db	4720	TGCTGCCGTAACAGTGAATTTAAAGATATCTAGAAGGTGAAGAGAGATCTATAAGATAT	4779
Qy	5457	CTTAGAAAAAGAAATAGTTTGAATAGAGCTTCTCTGATGACTATAGTGTGCTCTGATCAA	5516
Db	4780	CCAAGAGAATCTAACAATTAATAACCTTCCAGAGGATTTACAGCACTCATTAATCAA	4839
Qy	5517	GCTTCTCATTTTCAGTGTCCCGCTGTGCAGATGATGAGCAAAAGCACTCTGTCTCTGTC	5576
Db	4840	GCATCCAAATTTCTGCTGCCGGAATCAGGTGTTGATAAGAGCAGAGCCCCAACTCTGTGC	4899
Qy	5577	CTTTTCTGTGGGGTATATCTTCTCAGAACATTTTGTCTGCCAGGAAATTTGTGAACGGG	5636
Db	4900	CTTGTGTGCGGATCTCTGCTGTCTCCAGAGTTACTGCTGCCAGACTGAATGGAAGGG	4959
Qy	5637	GAGAGGTTGGAGTTGCAATTTTTCAGCACTTTCATCTGTGGAGCGGAGTCTGCAATTTTC	5696
Db	4960	GAGGATAGGAGCTGCACAGCTCACACCTTACTCTCTGTGCTCTGGAGTGGGCACTTTC	5019
Qy	5697	CTAAAAATCAGAGAAATGCCAGTGTCTGCTGTTGAAGTAAAGCCAGAGGCTGTGCCCTAT	5756
Db	5020	CTGAGAGTACGGGAATGTCAAGTGTCTATTTTGTGCAAAACCAAGGCTGTTTTTAT	5079
Qy	5757	CCAGCTCTTACTTGGATGAATATGGAGAAACAGACCTTGGCTTGAAGAGGGCAACCCC	5816
Db	5080	TCCTCTCTTACCTTGATGACTATGGGAGACCCACAGGAGCTCAGACGGGGAATCT	5139
Qy	5817	CTTCAATTATCTCGTAGCGGTATCGGAAGCTTCCATTTGGTCTGCAACAACTGCAAT	5876
Db	5140	TTACATTTATGCAAGAGCGATTTCAAGAAGATTTCAGAAGCTCTGGCAACACAGAGTCTC	5199
Qy	5877	ATAGAGAGATTCGTAGAGCCAGAGACTAATCAGATGTTTATTTGGATTTCACTGGCA	5935
Db	5200	ACAGAGGAAATTTGGACATGTCACAGAGAACCAATCAGACACTGGTTGGATTTGACTGGCA	5258

[illegible]

QY	1903	ATATCACTGATGATCATACAGAGAATATCTCTATAACTGCACCTTTCAGTTTCAGATGTTTA	1962
Db	1208	TTATGGAGGATGATCACAGCGGGCAGTGCCTGCTCTCTGCTCCAGCTTCTTCA	1267
QY	1963	CTGTTCCTACTCTGGCTCGACATCTTATTGAAGAGCAGATGTTATCTCTGTCTATTACTG	2022
Db	1268	CCGACCGACGCTGGCGGGAATGCTCCTCACAGAGAACTGTATGACCGTTATCATTA	1327
QY	2023	AAACTCTGCTAGAAATTTT-----ACCTGAGTACTTTGGACAGAAACAATAAATCAACT	2076
Db	1328	AGGCTTTTCTGACCAATTTTAAACACACAGAGATGCCAGGCGAGATTCAGCTTTGAACGCT	1387
QY	2077	TCCAGGGTTATAGCCAGGACAAATTGGGAAGAGTATATGCAGTAAATATGTGACCTAAAGT	2136
Db	1388	ACACTGCCCTCCAAAGCCCTTCAGTTCAGGAGAGTCCAGAGCCCTCATCTTAGATCTCAAGT	1447
QY	2137	ATATCCTTGATCAGCAAAACCCACAATATGGACAGAAAGATTTAAGAATGCAGTTCCTTTGAAG	2196
Db	1448	ATGTATTGATTAGCAAAACCAACGAGTGGTCAGATGAGCTGAGCGAGAAGTTCCTACAG	1507
QY	2197	GTTTTTCGATCTTTTTTGAAGATTCTTACCTGTATGAGGGAATGGAAGAAATCCGAAAGAC	2256
Db	1508	GGTTTCGATGCTCTCTTGGAAATTACTGAAGTGCATGCAGGGAATGGACCCGATCACGCGTC	1567
QY	2257	AGGTTGGGCAACACATTTGAAGTGCATCTCATTTGGGAGCTGCCAATGTCTATACAGATGC	2316
Db	1568	AGGTGGCAGACACATTTGAGATGAGCGCAGTGGGAAGCAGCGCTTCACACTGCAGATGA	1627
QY	2317	AATTTGAAGAATATTTTACTCATGTTCCAAGTGGTGTGCTGTGTATGAAGAACTCTTAC	2376
Db	1628	AGCTGACACACGTCATCTCAATGGTCAGCACCTGGTGTCTCTGGACGAAAAGTGTAA	1687
QY	2377	TTGTGGCTTATAAGAATGTACAAAGCTGTGATGAGGTGCATACACAGTTTCATATCTA	2436
Db	1688	TTGAAGCTTACAAGAAATGCTGGCTGTGTGCACAGATGCTATGGCGGATTTTACTGATG	1747
QY	2437	GTAGCAAGACAGT---AGTACAATCGGTGGACATAGTTTGGAAACAAAGTCCCTACAGAG	2493
Db	1748	GTGACAGCCAAATCACACTCAGTATTTTGTGGACACTCGGTGGAAACCATCAGATCTGTG	1807
QY	2494	TATCTGAGGATCTTGTAAGCATACATCTGCCACTCTCTAGGACCCTTGTGCTTTCATG	2553
Db	1808	TTTCCCAGAAAAAGTTAGCATTCACCTCCCAATTTCTCGCTTGTTCAGGTTTGCATG	1867
QY	2554	TACGTTTAAAGCAGGCTGGTGTGTTTCAAGACTGCATGAATTTGTGTCTTTTGGAGACT	2613
Db	1868	TATTGTTTAAAGCAAAAGTGAAGTGGCATATAAATTTCCAGAGCTCCACCTCTAAAGTGAAC	1927
QY	2614	TTCAAGTAGAGGTACTAGTGGAAATATCCTTTACGTTGTCTGTGGTGGTTGCCCGAGTTG	2673
Db	1928	TGAGCCCAACCCATGTTGATAGAACATCCTCTTAGATGTCGTGTCTTATGTGCTCAAGTGC	1987
QY	2674	TTGCTGAGATGTGGCGAAGAAATGGAGTGTCTCTTATTAGCCAGGTGTTTTATTATCCAAAG	2733
Db	1988	ATGCTGGATGTGGAGAAGAAATGGCTTCTCTCTAGTAAATCAGATCTATTACTACCAAT	2047
QY	2734	ATGTTAAGTGCAGAGAAGAAATGTATGATAAACATATCATCATGCTTTCAGATGGTGCAT	2793
Db	2048	ATGTGAATACGCGCGAGAGATTTCACAAAGACATAGTGTATGCTTCAGACAGGTGTCT	2107
QY	2794	CTTTAATGGATCCCAATAAGTCTCTGTTACTGTGTTCTCAGAGGTATGAACTTGCCGAGG	2853
Db	2108	CCATATGGACCCAAACCCATCTCCTGATGATCATGCTCAGCCGCTTTGAACCTCTATCAGC	2167
QY	2854	CTTTTAAACAAGACCATATCTTACAAAAG-----ACCAAGATTTGA	2892
Db	2168	TCCTCAGCAGCCTGCATATGGGAAGAGATTCAGTTCTGAGGTTACCCATAAGACGCTCG	2227
QY	2893	TTAAACAATATATACACTAATATAGAGAAATGCTTCAGGTCTCATCTATATTGTGGGTG	2952
Db	2228	TTCAGCAGAAACACACTCTGATCGAAGAGATGCTTACCTCATCATCTGCTTGTGGGAG	2287

QY	2953	AGCGTTATGTACCTGGAGTGGGAATGTGACCAAGAAAGAGGTGCACAAATGAGAGAAATCA	3011
Db	2288	AAAAGATTCAACCTTGGGGTTGGACAGGTGGCTGCCACAGATGAAATCAAGAGGGAGATTGA	2347
QY	3013	TTCACTTGTCTTGGCATTTGAACCCATGCCACACAGTGCATTGGCCAAAAATTTACCTTGAGA	3072
Db	2348	TCCATCAGTTGAGCATCAAGCCTATGGCTCACAGTGAGCTGGTGAAGTCTCTGCCCTGAAG	2407
QY	3073	ATGAAAATAATGAAACTGGCTTAGAGAAATGTCATAAACAAGTGGCCACATTTTAAGAAAC	3132
Db	2408	ATGAGAACAAAGGAGACCGCATGGAGAGCGTCATCGAGTCCGTTGCACATTTCAAGAAAC	2467
QY	3133	CAGGTGTATCAGGCCATGGAGTTTATGAACTAAAGATGAATCACTCACTGAAGACTTCAATA	3192
Db	2468	CTGGCTCAcAGGGCGAGGCATGTATGAGCTGAAGCCAGAGTGTGCCAAAGATTCAACC	2527
QY	3193	TGTACTTTTATCATTTACTCCAAAACCCAGCATAGCAAGGCTGAACATATGCAAGAAGAAA	3252
Db	2528	TGTATTTTATCATTTCTCCAGGCAGAGCAGTCCAAAGGCAGAGGAGCTCAGCGGAAAT	2587
QY	3253	GGAGAAACAAGAAACAAAGATGAAGCATTTGCCGCCACCACCACTCTCTGAAATTTGTC	3312
Db	2588	TGAAAAGAGAAAATAAAGAAAGATACAGCACTCCCTCCTCCGGCTTTGCCACCGTTCTGCC	2647
QY	3313	CTGCTTTCAGCAAGTGAATTAACCTTCTCAACTGTGATATCATGATGTACATCTTCAGGA	3372
Db	2648	CTTTGTTGGGAGTCTGGTTTAACATCTTCAGATGTGACGTGATGCTGTACATCATGGGA	2707
QY	3373	CCGTATTTGAGCGGGCAATAGACACAGATTTCTAACTGTGCGACCGAAGGATGCTCCAAA	3432
Db	2708	CGATCTCGAGTGGGCTGTAGAGCATCACGGGTCTGCGCTGGTCAGAGTCCATGCTACAGA	2767
QY	3433	TGGCTTTTTCATATCTGGCATTTGGGTTTACTAGAAGAGAAAGCAACAGCTTCAAAAAGCTC	3492
Db	2768	GGGTGCTGCATTTGATCGGATGSGCTCTCCAGGAAGAGAACCACTTGGAGAACCGCG	2827
QY	3493	CTGAAGAAGAAGT---AACATTTGACHTTTATCATTAAGGCTTCAAGATTTGGGAAGTTTCAG	3549
Db	2828	TGGAAGGGCAGTGCAGACCTTCACTTACACAGAAAGATTTCAAAGCCTTGGTGTATGCAC	2887
QY	3550	-----CCATGAATATACAATGCTTTTGGAAAAAATCAAGGAATTTCCCCAGTTAG	3600
Db	2888	CACATAACTCCCGAGCATCCTAGCTATGCTGGAGACCTTGCAGACGCCCTCCCTGG	2947
QY	3601	AAGCCCAAGAGCATGATAACGTGGATCTTCAGATGTTTGCACAGTGAAGCGATTA	3660
Db	2948	AAGCCCAAGAGCATGATCAGGTGGTGTCTAAAGATGTTTAATGCAATTAAGAAGATA	3007
QY	3661	GAGAAAATCTGTTTAATTTGTAGCAACACATCAGGATCGGAATCTATTAAAGATGATG	3720
Db	3008	GAGAGTCTTCATCCAGCAGCCCTGTGGCC-----GAGGCGGAGGGAACCATATGGAGG	3061
QY	3721	AGATTACTGATTAAGAAAAGCAGAACGAAAAGAAAGCTGAAGCTGTAGCTAGCTAC	3780
Db	3062	AGAGCTCAAGAGACAAGGCAAGACAGAGAGGAAAAGAAAGCCGAGATGCGCAGACTGC	3121
QY	3781	ATCCCCAGAAGATCATGGCTCAGATGCTGCCTTACAGAAAATCTCATTTGAAATCATA	3840
Db	3122	GCCGGGAGAGATCATGCCCCAGATGCTCAGATGCAAGCGGCATCTATTGACGAAAACA	3181
QY	3841	AACCTCATGTATGACAAATPACATCAGAAATGCTTGGGAAAGAAAGATTCATATTGGAGGAAG	3900
Db	3182	AAGAGCTCTCCAGCAGACCCTAGAGCTGGACACCTCTGCTCTGCGCACTCT-----TG	3235
QY	3901	AGAGCACCACAGAGTCAGTACTCTPAGAATTTGCTTTGGTCTCTAAACGGGTCCAT	3960
Db	3236	ACAGCAGCCCTCCGTTTTCAGAGCGAGCTCTTACAGCACTTGGGCCACAGACACAGG	3295
QY	3961	CTGTTACTGAAAAGGAGGTGCTGACGTGCATCTTTGCCAAGAAAGAACAGGAGGTCAAAA	4020
Db	3296	TCCCTGAACCGAGACAGTTTGTCCACTGTATATATGTCAAGAGGAGCAAGAGGTGACTG	3355
QY	4021	TAGAAAATAATGCCATGGTATTATPCCGCTGTGTCCAGAAATCTACTGCCCTTAACCCAGC	4080

Db	3356	I I
----	------	---

FEATURES
source Location/Qualifiers
1..1001
/organism="unknown"
BASE COUNT 363 a 186 c 205 g 247 t
ORIGIN

Query Match 15.8%; Score 996.2; DB 6; Length 1001;
Best Local Similarity 99.7%; Pred. No. 9.4e-241;
Matches 998; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2799 ATGGATCCCAATAAGTCTTGTACTGGTACTTACAGAGGTATGAACCTTGCAGGCTTTT 2858
Db 1 ATGGATCCCAACAAGTCTTGTACTGGTACTTACAGAGGTATGAACCTTGCAGGCTTTT 60
QY 2859 AACAAGACCATATCTCAAAAGACAGGATTTGATTAAACAATATATACACTAATAGAA 2918
Db 61 AACAAGACCATATCTCAAAAGACAGGATTTGATTAAACAATATATACACTAATAGAA 120
QY 2919 GAAATGCTTCAGTCTCATCTATATTGTGGTGAGGTTATGTACCTGGAGTGGGAAT 2978
Db 121 GAAATGCTTCAGTCTCATCTATATTGTGGTGAGGTTATGTACCTGGAGTGGGAAT 180
QY 2979 GTGACCAAGAAGAGGTCACAATGAGAGAAATCATTACATTTGCTTTGCATTGAACCCATG 3038
Db 181 GTGACCAAGAAGAGGTCACAATGAGAGAAATCATTACATTTGCTTTGCATTGAACCCATG 240
QY 3039 CCACACAGTGCCTATGCCAAAATTTACCTGAGAAATTAATGAACCTGGCTTAGAG 3098
Db 241 CCACACAGTGCCTATGCCAAAATTTACCTGAGAAATTAATGAACCTGGCTTAGAG 300
QY 3099 AATGTCATRAACAAGTGGCCACATTTAAGAAACCAGGTGTATCAGGCCATGGAGTTTAT 3158
Db 301 AATGTCATRAACAAGTGGCCACATTTAAGAAACCAGGTGTATCAGGCCATGGAGTTTAT 360
QY 3159 GAACTAAAAGATCAATCACTGAAAGACTTCAATATGTACTTTTATCATTTACTCCAAAACC 3218
Db 361 GAACTAAAAGATCAATCACTGAAAGACTTCAATATGTACTTTTATCATTTACTCCAAAACC 420
QY 3219 CAGCATAGCAAGGCTGAACATATGCAGAGAAAGAGAGAAACAAGAGAAACAAGATGAA 3278
Db 421 CAGCATAGCAAGGCTGAACATATGCAGAGAAAGAGAGAAACAAGAGATGAA 480
QY 3279 GCATTCGGCCACACACCTCTCTGAATTCCTGCCCTGCTTTCAGCAAAAGTATTAACCTT 3338
Db 481 GCATTCGGCCACACACCTCTCTGAATTCCTGCCCTGCTTTCAGCAAAAGTATTAACCTT 540
QY 3339 CTCACACTGTGATATCATGTATCTCAGGACCGTATTTGAGCGGGCAATAGACACA 3398
Db 541 CTCACACTGTGATATCATGTATCTCAGGACCGTATTTGAGCGGGCAATAAACACA 600
QY 3399 GATTCCTAACTTGTGGACCGAGGATGCTCCAAATGGCTTTTTCATATTCCTGGCATTTGGT 3458
Db 601 GATTCCTAACTTGTGGACCGAGGATGCTCCAAATGGCTTTTTCATATTCCTGGCATTTGGT 660
QY 3459 TTACTAGAAGAGAGCAACAGCTTCAAAAAGCTCCTGAAGAAGAGTAACATTTTGACTTT 3518
Db 661 TTACTAGAAGAGAGCAACAGCTTCAAAAAGCTCCTGAAGAAGAGTAACATTTTGACTTT 720
QY 3519 TATCATAAAGCTTCAAGATTTGGGAAGTTAGCCATGAATATACAAATGCTTTTGGAAAAA 3578
Db 721 TATCATAAAGCTTCAAGATTTGGGAAGTTAGCCATGAATATACAAATGCTTTTGGAAAAA 780
QY 3579 CTCAAAGGAATTCCTCCAGTTAGAGCCCAAGGACATGATAACCTGGATCTTCAGATG 3638
Db 781 CTCAAAGGAATTCCTCCAGTTAGAGCCCAAGGACATGATAACCTGGATCTTCAGATG 840
QY 3639 TTTGACACAGTGAAGCGATTGAAGAAAAATCTTGTGTTAATTTGTAGCAACCATCATCAGGA 3698
Db 841 TTTGACACAGTGAAGCGATTGAAGAAAAATCTTGTGTTAATTTGTAGCAACCATCATCAGGA 900
QY 3699 TCGGAATCTATTGAAGATGATGAGATTTACTCATGATAAAGAAAAAGCAGAACCAAGAAAGA 3758
Db 901 TCGGAATCTATTGAAGATGATGAGATTTACTCATGATAAAGAAAAAGCAGAACCAAGAAAGA 960

QY 3759 AAAGCTGAAGCTGTAGCTACATCCCGAAGATCATGGC 3799
Db 961 AAAGCTGAAGCTGTAGCTTTCATCCCGAAGATCATGGC 1001
RESULT 13
ARI21464
LOCUS ARI21464 1001 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 2 from patent US 6159732.
ACCESSION ARI21464
VERSION ARI21464.1 GI:14105040
KEYWORDS
SOURCE unknown.
ORGANISM unknown.
REFERENCE 1 (bases 1 to 1001)
AUTHORS Varshavsky, A. and Kwon, Y. Tae.
TITLE Nucleic acid encoding mammalian Ubr1
JOURNAL Patent: US 6159732-A 2 12-DEC-2000;
FEATURES Location/Qualifiers
1..1001
source /organism="unknown"
BASE COUNT 363 a 186 c 205 g 247 t
ORIGIN
Query Match 15.8%; Score 996.2; DB 6; Length 1001;
Best Local Similarity 99.7%; Pred. No. 9.4e-241;
Matches 998; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2799 ATGGATCCCAATAAGTCTTGTACTGGTACTTACAGAGGTATGAACCTTGCAGGCTTTT 2858
Db 1 ATGGATCCCAACAAGTCTTGTACTGGTACTTACAGAGGTATGAACCTTGCAGGCTTTT 60
QY 2859 AACAGAGCCATATCTACAAAAGACAGGATTTGATTAAACAATATATACACTAATAGAA 2918
Db 61 AACAGAGCCATATCTACAAAAGACAGGATTTGATTAAACAATATATACACTAATAGAA 120
QY 2919 GAAATGCTTCAGGCTCCTCATCTATATTGTGGTGAGCGTTATGTACTCGAGTGGGAAAT 2978
Db 121 GAAATGCTTCAGGCTCCTCATCTATATTGTGGTGAGCGTTATGTACTCGAGTGGGAAAT 180
QY 2979 GTGCCAAAGAAGAGGTCACAATGAGAGAAATCATTCACCTTGTGCTATGAACCCATG 3038
Db 181 GTGCCAAAGAAGAGGTCACAATGAGAGAAATCATTCACCTTGTGCTATGAACCCATG 240
QY 3039 CCACACAGTGCCTATGCCAAAATTTACCTGAGAAATTAATGAACCTGGCTTAGAG 3098
Db 241 CCACACAGTGCCTATGCCAAAATTTACCTGAGAAATTAATGAACCTGGCTTAGAG 300
QY 3099 AATGTCATRAACAAGTGGCCACATTTAAGAAACCAGGTGTATCAGGCCATGGAGTTTAT 3158
Db 301 AATGTCATRAACAAGTGGCCACATTTAAGAAACCAGGTGTATCAGGCCATGGAGTTTAT 360
QY 3159 GAACTAAAAGATGAATCACTGAAAGACTTCAATATGTACTTTTATCATTTACTCCAAAACC 3218
Db 361 GAACTAAAAGATGAATCACTGAAAGACTTCAATATGTACTTTTATCATTTACTCCAAAACC 420
QY 3219 CAGCATAGCAAGGCTGAACATATGCAGAGAAAGAGAGAAACAAGAGAAACAAGATGAA 3278
Db 421 CAGCATAGCAAGGCTGAACATATGCAGAGAAAGAGAGAAACAAGAGATGAA 480
QY 3279 GCATTCGGCCACACACCTCTCTGAATTCCTGCCCTGCTTTCAGCAAAAGTATTAACCTT 3338
Db 481 GCATTCGGCCACACACCTCTCTGAATTCCTGCCCTGCTTTCAGCAAAAGTATTAACCTT 540
QY 3339 CTCACACTGTGATATCATGTATCTCAGGACCGTATTTGAGCGGGCAATAGACACA 3398
Db 541 CTCACACTGTGATATCATGTATCTCAGGACCGTATTTGAGCGGGCAATAAACACA 600
QY 3399 GATTCCTAACTTGTGGACCGAGGATGCTCCAAATGGCTTTTTCATATTCCTGGCATTTGGT 3458
Db 601 GATTCCTAACTTGTGGACCGAGGATGCTCCAAATGGCTTTTTCATATTCCTGGCATTTGGT 660
QY 3459 TTACTAGAAGAGAGCAACAGCTTCAAAAAGCTCCTGAAGAAGAGTAACATTTTGACTTT 3518
Db 661 TTACTAGAAGAGAGCAACAGCTTCAAAAAGCTCCTGAAGAAGAGTAACATTTTGACTTT 720
QY 3519 TATCATAAAGCTTCAAGATTTGGGAAGTTAGCCATGAATATACAAATGCTTTTGGAAAAA 3578
Db 721 TATCATAAAGCTTCAAGATTTGGGAAGTTAGCCATGAATATACAAATGCTTTTGGAAAAA 780
QY 3579 CTCAAAGGAATTCCTCCAGTTAGAGCCCAAGGACATGATAACCTGGATCTTCAGATG 3638
Db 781 CTCAAAGGAATTCCTCCAGTTAGAGCCCAAGGACATGATAACCTGGATCTTCAGATG 840
QY 3639 TTTGACACAGTGAAGCGATTGAAGAAAAATCTTGTGTTAATTTGTAGCAACCATCATCAGGA 3698
Db 841 TTTGACACAGTGAAGCGATTGAAGAAAAATCTTGTGTTAATTTGTAGCAACCATCATCAGGA 900
QY 3699 TCGGAATCTATTGAAGATGATGAGATTTACTCATGATAAAGAAAAAGCAGAACCAAGAAAGA 3758
Db 901 TCGGAATCTATTGAAGATGATGAGATTTACTCATGATAAAGAAAAAGCAGAACCAAGAAAGA 960

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2003, 19:32:57 : Search time 993.895 Seconds
(without alignments)
17132.652 Million cell updates/sec

Title: US-09-724-126A-1
Perfect score: 6308
Sequence: 1 gccagaattcgccacgagg.....aattttgtatttggtgtttt 6308

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03.*
1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
25: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB ID	Description
1	5592	88.6	7742	25	ABZ24689	Human cell growth,
2	4435.4	70.3	6395	20	AAV99308	cDNA encoding mous
3	4435.4	70.3	6395	22	AAV99308	Nucleotide sequenc
4	2546.8	40.4	2550	22	AAH14878	Human cDNA sequenc
5	1311	20.8	1635	24	ABQ75898	Human ubiquitin re
6	1099.4	17.4	6840	22	AAK51709	Human polynucleoti
7	1090	17.3	6850	22	AAK52693	Human polynucleoti
8	996.2	15.8	1001	20	AAV03300	Partial cDNA encod

9	996.2	15.8	1001	22	AAV86934	Nucleotide sequenc
10	813.6	12.9	3327	20	AAV35730	cDNA encoding a pr
11	813.6	12.9	3502	20	AAV35731	Human colon cancer
12	656.8	10.4	733	21	AAH02411	Human cDNA clone (
13	647	10.3	712	22	AAH07621	Human colon cancer
14	638.6	10.1	756	21	AAV02327	Fibrinogen 9,57 co
15	626.8	9.9	2616	24	ABZ70161	DNA encoding novel
16	450.8	7.1	777	23	AAV70068	DNA encoding novel
17	429	6.8	1464	23	AAV70104	Human polynucleoti
18	374.8	5.9	1772	24	ABL90577	DNA encoding novel
19	271.4	4.3	1172	23	AAV70851	Human cDNA encodin
20	250.6	4.0	972	24	ABK35394	Gastric cancer ass
21	246.8	3.9	807	20	AAV39891	Human colon cancer
22	241.6	3.8	264	21	AAV00233	cDNA upregulated i
23	219.8	3.5	4573	25	ACA03886	Drosophila melanog
24	216.8	3.4	6278	23	ABL14799	Gastric cancer ass
25	191.6	3.0	800	20	AAV39892	EST clone DA490.
26	159	2.5	455	20	AAV89674	Human polynucleoti
27	142	2.3	643	24	ABL89546	Drosophila melanog
28	132	2.1	13134	23	ABL14798	Human immune/haema
29	128	2.0	529	22	AAV62305	Human gene signatu
30	116.6	1.8	308	16	AAV21876	Human secreted pro
31	105.6	1.7	366	21	AAV00938	Human secreted pro
32	105.6	1.7	456	21	AAV00940	Human ovarian anti
33	105.6	1.7	1459	24	ABQ54866	Human cancer assoc
34	105.6	1.7	1461	21	AAV78096	Human prostate exp
35	104.8	1.7	507	23	ABV59104	Human novel polynu
36	104.8	1.7	1401	25	ABV05003	Human polynucleoti
37	103.2	1.6	428	22	AAI91546	Human secreted pro
38	103.2	1.6	605	21	AAV00939	Human secreted pro
39	100.8	1.6	602	21	AAV10733	Human secreted pro
40	100.8	1.6	692	21	AAV10734	Human cDNA for nov
41	100.8	1.6	697	24	ABK34403	Human secreted pro
42	100.8	1.6	842	21	AAV10872	Human cDNA differe
43	100.8	1.6	1370	24	ABK83603	Human lung specifi
44	100.6	1.6	1382	25	ABQ82975	Bovine EST associa
45	89.6	1.4	326	25	ABX49787	

ALIGNMENTS

RESULT 1	
ABZ24689	
ID	ABZ24689 standard; cDNA; 7742 BP.
XX	
AC	ABZ24689;
XX	
DT	07-APR-2003 (first entry)
XX	
DE	Human cell growth, differentiation and death protein CGDD-1 cDNA.
XX	
KW	CGDD-1; cell growth; cell differentiation; cell death; human;
KW	cytosolic; antiarteriosclerotic; hepatotropic; antiinflammatory;
KW	antiproliferative; antianemic; ophthalmological; auditory;
KW	anticonvulsant; cerebroprotective; nootropic; neuroprotective;
KW	antiparkinsonian; neuroleptic; tranquilizer; immunosuppressive;
KW	anti-HIV; antiallergic; antiasthmatic; antithyroid; antidiabetic;
KW	dermatological; nephrotropic; antirheumatic; antiarthritic;
KW	antitumor; vulnary; virucide; antibacterial; fungicide;
KW	antiparasitic; protozoacide; antihelminthic; antiinfertility;
KW	gynaecological; ubiquitin protein ligase; enzyme; gene therapy;
KW	microarray; gene; ss.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	CDS
FT	Location/Qualifiers
FT	13..5229
FT	/*tag= a
FT	/product= "Human CGDD-1"
XX	
PN	WO200297032-A2.
XX	

PD 05-DEC-2002.
 XX
 PF 05-APR-2002; 2002WO-US11152.
 XX
 PR 06-APR-2001; 2001US-282110P.
 PR 11-APR-2001; 2001US-283294P.
 PR 26-APR-2001; 2001US-286820P.
 PR 27-APR-2001; 2001US-287228P.
 PR 16-MAY-2001; 2001US-291662P.
 PR 18-MAY-2001; 2001US-291846P.
 PR 25-MAY-2001; 2001US-293727P.
 PR 01-JUN-2001; 2001US-295263P.
 PR 01-JUN-2001; 2001US-295340P.
 PR 15-JAN-2002; 2002US-349705P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Azimzai Y, Au-Young JK, Batra S, Baughn MR, Becha SD, Borowsky ML;
 PI Burford N, Ding L, Elliott VS, Emerling BM, Gandhi AR, Gietzen KJ;
 PI Griffin JA, Hafalia AJA, Honchell CD, Lal PG, Lee SY, Lu DAM;
 PI Arvizu CS, Ramkumar J, Reddy R, Sanjanwala MM, Tang YT, Wallia NK;
 PI Wang YE, Warren BA, Xu Y, Yang J, Yao MG, Yue H, Zebbarjadian Y;
 XX
 DR WPI; 2003-140453/13.
 DR P-PSDB; ABP58330.
 XX
 PT Novel human proteins associated with cell growth, differentiation and
 PT death, useful for treating, diagnosing or preventing cancer,
 PT developmental, neurological, reproductive or autoimmune/inflammatory
 PT disorders -
 XX
 PS Claim 5; Page 216-218; 238pp; English.
 XX
 CC The present sequence is that of Incyte clone 1351608CB1 encoding
 CC human CGDD-1, a novel protein associated with cell growth, for the
 CC differentiation and death. A representative cDNA library for the
 CC polynucleotide is PGANN0701 from paraganolionic tumour tissue.
 CC Structural features establish the encoded protein as being
 CC associated with cell growth, differentiation and death, with
 CC further evidence suggesting it to be a ubiquitin protein ligase.
 CC The invention is based on novel human CGDD-1 to -21 proteins (see
 CC ABP58330-507), the polynucleotides encoding them (see AB224689-709),
 CC and to the use of these for the diagnosis, treatment or prevention
 CC of cell proliferative disorders including cancer, developmental
 CC disorders, neurological disorders, autoimmune disorders,
 CC reproductive disorders, and disorders of the placenta, and in the
 CC assessment of the effects of exogenous compounds on the activity
 CC and expression of proteins and nucleic acids associated with cell
 CC growth, differentiation and death. CGDD polynucleotides are also
 CC used in a claimed microarray and in a claimed method of generating
 CC an expression profile of a sample.
 XX
 SX Sequence 7742 BP; 2388 A; 1423 C; 1622 G; 2309 T; 0 other;
 SX
 Query Match 88.6%; Score 5592; DB 25; Length 7742;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 5592; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 717 GGTAAGAGAGATGGAATACAGCGGGAGTATACCCAGACCCCTCAGCGTCTGGCATCT 776
 DB 1 GGTAAGAGAGATGGAATACAGCGGGAGTATACCCAGACCCCTCAGCGTCTGGCATCT 60
 QY 777 TGGTGGATCAGCAAGTGGATTTTATACATGCTTTCTTGGATCATTTGGCACAATTTGGT 836
 DB 61 TGGTGGATCAGCAAGTGGATTTTATACATGCTTTCTTGGATCATTTGGCACAATTTGGT 120
 QY 837 CCAGAAATTTACATTTGCTGAAATGGACCCAGACTTGGAAAAGCAGGAGGAAAGTGTACAA 896
 DB 121 CCAGAAATTTACATTTGCTGAAATGGACCCAGACTTGGAAAAGCAGGAGGAAAGTGTACAA 180
 QY 897 ATGTCAATATTCACCTCCAGTGAATGGTACTTATTTGGAGAAGATCCAGATATTTGGCTTA 956
 DB 181 ATGTCAATATTCACCTCCAGTGAATGGTACTTATTTGGAGAAGATCCAGATATTTGGCTTA 240

QY 957 GAGAAATTGAAGACAGGTGGAGCATTTTCAGCTTTTGTGGAGGGTTTTCAAAAGTGGAGAG 1016
 DB 241 GAGAAATTGAAGACAGGTGGAGCATTTTCAGCTTTTGTGGAGGGTTTTCAAAAGTGGAGAG 300
 QY 1017 ACAACCTATTCTTGCAGGGATTGTGCAATTGATCAACATGTGTACTCTGTATGACTGCG 1076
 DB 301 ACAACCTATTCTTGCAGGGATTGTGCAATTGATCAACATGTGTACTCTGTATGACTGCG 360
 QY 1077 TTCCAGGACAGTGTTCATAAAAAATCATGTTACAAGATGCATACCTTCTACTCGAGGAGG 1136
 DB 361 TTCCAGGACAGTGTTCATAAAAAATCATGTTACAAGATGCATACCTTCTACTCGAGGAGG 420
 QY 1137 TTCTGTGACTGTGGAGACACAGAGGCATGGAAAACTGGCCCCCTTTTGTGTAAATCATGAA 1196
 DB 421 TTCTGTGACTGTGGAGACACAGAGGCATGGAAAACTGGCCCCCTTTTGTGTAAATCATGAA 480
 QY 1197 CCTGGAAGACAGGTACTATTAAGAGAAATTCACCTGTCCGTTGGAATGAAGAGTAATT 1256
 DB 481 CCTGGAAGACAGGTACTATTAAGAGAAATTCACCTGTCCGTTGGAATGAAGAGTAATT 540
 QY 1257 GTCCAAGCCAGAAAAATTTTCCCTTCACTGATATAAATATGTCGTAGAAATGACTATATGG 1316
 DB 541 GTCCAAGCCAGAAAAATTTTCCCTTCACTGATATAAATATGTCGTAGAAATGACTATATGG 600
 QY 1317 GAAGAGAAAAAGAACTGCCTCCTGAACCTCCAGATAAGGGAGAAAAATGAAAGATACTAT 1376
 DB 601 GAAGAGAAAAAGAACTGCCTCCTGAACCTCCAGATAAGGGAGAAAAATGAAAGATACTAT 660
 QY 1377 TGTGTCTCTTTCAATGATGAACACCATTTCAATATGACACGTCATATACAGCCTCAAGA 1436
 DB 661 TGTGTCTCTTTCAATGATGAACACCATTTCAATATGACACGTCATATACAGCCTCAAGA 720
 QY 1437 GCTCTTGACTGTGAGCTCGCAGAGCCAGTTCATACCCAGTCCACTGCAAGAAAGATATAAG 1496
 DB 721 GCTCTTGACTGTGAGCTCGCAGAGCCAGTTCATACCCAGTTCGCAAGAAAGATATAAG 780
 QY 1497 CGTCGGGCTGTTAAAGCGGGAGCTTATGCTGCTGCCAGGAAGCAAGAAAGATATAAG 1556
 DB 781 CGTCGGGCTGTTAAAGCGGGAGCTTATGCTGCTGCCAGGAAGCAAGAAAGATATAAG 840
 QY 1557 AGTCATTCAGAAAAATGCTCTCAACATCCATTCATGATAGAAAGTATACACATCAGAGATT 1616
 DB 841 AGTCATTCAGAAAAATGCTCTCAACATCCATTCATGATAGAAAGTATACACATCAGAGATT 900
 QY 1617 ATGCTCATCAGAAATTTGCTTGGCTCTGCTGCTGATCAACAAATATGAGCTAT 1676
 DB 901 ATGCTCATCAGAAATTTGCTTGGCTCTGCTGCTGATCAACAAATATGAGCTAT 960
 QY 1677 TCAAGTGAATTTAGGCAGATCTTTTCCCAAGCATGCTTTAGAGAAAGCACTGACCTCGGAG 1736
 DB 961 TCAAGTGAATTTAGGCAGATCTTTTCCCAAGCATGCTTTAGAGAAAGCACTGACCTCGGAG 1020
 QY 1737 AATCCCTGCTCATAGCAGGTTAATGCTTTGGGATGCAAGCTTTTAAAGGTGCCCT 1796
 DB 1021 AATCCCTGCTCATAGCAGGTTAATGCTTTGGGATGCAAGCTTTTAAAGGTGCCCT 1080
 QY 1797 AAGATCCTTTCATGAATTTGATCTTACAGCAGTTTTTTATGAGAGATGGAATACAAAACATC 1856
 DB 1081 AAGATCCTTTCATGAATTTGATCTTACAGCAGTTTTTTATGAGAGATGGAATACAAAACATC 1140
 QY 1857 TTTGCTATGGAATTTGTGAAGTATTATTAACAACTGCAGAAAGATATATCAGTGTATGAT 1916
 DB 1141 TTTGCTATGGAATTTGTGAAGTATTATTAACAACTGCAGAAAGATATATCAGTGTATGAT 1200
 QY 1917 CATGACAGAAGTATCTCTATAACCTGACCTTTTCAGTTCAGATGTTTACTGTTCTACTCTG 1976
 DB 1201 CATGACAGAAGTATCTCTATAACCTGACCTTTTCAGTTCAGATGTTTACTGTTCTACTCTG 1260
 QY 1977 GCTGACATCTTATGGAAGACAGAGAAATGTTATCTCTGTCATTTACTGAAACTCTGCTAGAA 2036
 DB 1261 GCTGACATCTTATGGAAGACAGAGAAATGTTATCTCTGTCATTTACTGAAACTCTGCTAGAA 1320

2037 GTTTTACCTGAGTACTTGGACAGGAACAATAAATCAACTTCCAGGGTTATAGCCAGGAC 2096
1321 GTTTTACCTGAGTACTTGGACAGGAACAATAAATCAACTTCCAGGGTTATAGCCAGGAC 1380
2097 AAATTTGGGAAGAGATATATGACGTAATATGACCTAAAGATATATCCTGATCAGCAAAACC 2156
1381 AAATTTGGGAAGAGATATATGACGTAATATGACCTAAAGATATATCCTGATCAGCAAAACC 1440
2157 ACAATATGGACAGAAAGATTAAGAAATCCAGTTCCTTGAAGGTTTTCGATCTTTTGAAG 2216
1441 ACAATATGGACAGAAAGATTAAGAAATCCAGTTCCTTGAAGGTTTTCGATCTTTTGAAG 1500
2217 ATTCTTACCTGTATGACAGGAATGGAAGAAATCCGAAGACAGGTTGGCAACACATTTGAA 2276
1501 ATTCTTACCTGTATGACAGGAATGGAAGAAATCCGAAGACAGGTTGGCAACACATTTGAA 1560
2277 GTGGATCTGATTTGGGAGGCTGCCATTTGCTATACAGATGCAATTTGAAGATATTTTACTC 2336
1561 GTGGATCTGATTTGGGAGGCTGCCATTTGCTATACAGATGCAATTTGAAGATATTTTACTC 1620
2337 ATGTTTCCAAGAGTGGTGTCTTGTGATGAAGAACTCTTACTTGTGGCTTATAAAGAATGT 2396
1621 ATGTTTCCAAGAGTGGTGTCTTGTGATGAAGAACTCTTACTTGTGGCTTATAAAGAATGT 1680
2397 CACAAAGCTGTGATGAGTGCAGTACCAGTTTTCATATCTAGTACGACACAGTAGTACAA 2456
1681 CACAAAGCTGTGATGAGTGCAGTACCAGTTTTCATATCTAGTACGACACAGTAGTACAA 1740
2457 TCCTGTGGACATAGTTTGGAAACAAAGTCCACAGAGTATCTGAGGATCTTTGAAGATA 2516
1741 TCCTGTGGACATAGTTTGGAAACAAAGTCCACAGAGTATCTGAGGATCTTTGAAGATA 1800
2517 CATCTGCCACTCTTAGGACCCCTTGCTGCTTCAATGCTAGTCTTAAAGCAGCTGGGTGCT 2576
1801 CATCTGCCACTCTTAGGACCCCTTGCTGCTTCAATGCTAGTCTTAAAGCAGCTGGGTGCT 1860
2577 GTTTCAGAGCTGATGAATTTGCTTTTGGAGACTTTCAGTAGAGGTACTAGTGGAA 2636
1861 GTTTCAGAGCTGATGAATTTGCTTTTGGAGACTTTCAGTAGAGGTACTAGTGGAA 1920
2637 TATCCCTTACCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2696
1921 TATCCCTTACCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1980
2697 GGACTGCTCTTATTAGCCAGGTGCTTTTATTACCAAGATGTTAAGTCAGAGAGAAATG 2756
1981 GGACTGCTCTTATTAGCCAGGTGCTTTTATTACCAAGATGTTAAGTCAGAGAGAAATG 2040
2757 TATGATAAAGATATCATCATGCTTTCAGATTTGGTGATCTTTAATGGATCCCAATGATTC 2816
2041 TATGATAAAGATATCATCATGCTTTCAGATTTGGTGATCTTTAATGGATCCCAATGATTC 2100
2817 TTCTTACTGTACTTCAGAGGTATGAATTTGCCAGGCTTTTACAGACCATATCTACA 2876
2101 TTCTTACTGTACTTCAGAGGTATGAATTTGCCAGGCTTTTACAGACCATATCTACA 2160
2877 AAAGACCAGGATTTGATTAACCAATATAATACACTAATAAGAAATGCTTCAGGTCCTC 2936
2161 AAAGACCAGGATTTGATTAACCAATATAATACACTAATAAGAAATGCTTCAGGTCCTC 2220
2937 ATCTATATTGTGGGTGAGCCTTATGTACCTGGAGTGGAAATGTGACCAAGAGAGGTC 2996
2221 ATCTATATTGTGGGTGAGCCTTATGTACCTGGAGTGGAAATGTGACCAAGAGAGGTC 2280
2997 ACAATGAGAGAAATCATCTACTGCTTTGCTTGAATGAACCCATGCCACACAGTGGCC 3056
2281 ACAATGAGAGAAATCATCTACTGCTTTGCTTGAATGAACCCATGCCACACAGTGGCC 2340
3057 AAAAAATTTACCTGAGATGAATAATGAACCTGGCTTACAGATGTTCATAAACAAGTG 3116
2341 AAAAAATTTACCTGAGATGAATAATGAACCTGGCTTACAGATGTTCATAAACAAGTG 2400
3117 GCCACATTTAAGAAACCAGGTGTATCAGGCCATGGAGTTTATGAACCTAAAAGATGAATCA 3176

2401 GCCACATTTAAGAAACCAGGTGTATCAGGCCATGGAGTTTATGAACCTAAAAGATGAATCA 2460
3177 CTGAAAGACTTCAATATATGTACTTTTATCATTTACTCCAAAACCCAGCATAGCAAGGTGAA 3236
2461 CTGAAAGACTTCAATATATGTACTTTTATCATTTACTCCAAAACCCAGCATAGCAAGGTGAA 2520
3237 CATATGCAGAGAAAGGAGAAACAAGAAACAAGATGAAGCATTTGCCGCCACCACCA 3296
2521 CATATGCAGAGAAAGGAGAAACAAGAAACAAGATGAAGCATTTGCCGCCACCACCA 2580
3297 CCTCTCGAATTCCTGCCCTTCTCAGCAAGTGAATTAACCTTCTCAACTGTGTATATCATG 3356
2581 CCTCTCGAATTCCTGCCCTTCTCAGCAAGTGAATTAACCTTCTCAACTGTGTATATCATG 2640
3357 ATGTACATTCCTCAGGACCGTATTTGAGCGGCAATAGACACAGATTTCAACTTGTGGACC 3416
2641 ATGTACATTCCTCAGGACCGTATTTGAGCGGCAATAGACACAGATTTCAACTTGTGGACC 2700
3417 GAAGGATGCTCCCAATGGCTTTTATATTTCTGGCATTTGGGTTTACTAGAGAGAGCA 3476
2701 GAAGGATGCTCCCAATGGCTTTTATATTTCTGGCATTTGGGTTTACTAGAGAGAGCA 2760
3477 CAGCTTCAAAAGCTCCTGAAGAAGTAACATTTTGAATTTTATCATAGGCTTCAAGA 3536
2761 CAGCTTCAAAAGCTCCTGAAGAAGTAACATTTTGAATTTTATCATAGGCTTCAAGA 2820
3537 TTGGAGTTCAGCCATGAATATACAAATGCTTTTGGAAAACTCAAGGAATTTCCCCAG 3596
2821 TTGGAGTTCAGCCATGAATATACAAATGCTTTTGGAAAACTCAAGGAATTTCCCCAG 2880
3597 TTAGAAGGCCAGAGACATGATACGCTGGATCTTTCAGATGTTTACACAGTGAAGCGA 3656
2881 TTAGAAGGCCAGAGACATGATACGCTGGATCTTTCAGATGTTTACACAGTGAAGCGA 2940
3657 TTAAGAGAAAAATCTGTTTAAATTTAGCAACCATCATCAGGATCGGAATCTATTAGAAT 3716
2941 TTAAGAGAAAAATCTGTTTAAATTTAGCAACCATCATCAGGATCGGAATCTATTAGAAT 3000
3717 GATGAGATTACTCATGATAAAGAAAGCAGACGAAAAAGCTGAAGCTGCTAGG 3776
3001 GATGAGATTACTCATGATAAAGAAAGCAGACGAAAAAGCTGAAGCTGCTAGG 3060
3777 CTACATGCCAGAGATCATGCTCAGATGCTGCTTACAGAAAACTTCAATGAAACT 3836
3061 CTACATGCCAGAGATCATGCTCAGATGCTGCTTACAGAAAACTTCAATGAAACT 3120
3837 CATAACTCATGTATGACAAATACATCAGAAATGCTTGGGAAGAAAGATTCATATGGAG 3896
3121 CATAACTCATGTATGACAAATACATCAGAAATGCTTGGGAAGAAAGATTCATATGGAG 3180
3897 GAAGAGACACCCAGCAGTCACTACTAGAAATGCTTGGGTCTCTAAACGGGT 3956
3181 GAAGAGACACCCAGCAGTCACTACTAGAAATGCTTGGGTCTCTAAACGGGT 3240
3957 CCATCTGTACTGAAAGAGGAGTGTGAGTGCATCTTGGCAAGAAAGAGAGGTG 4016
3241 CCATCTGTACTGAAAGAGGAGTGTGAGTGCATCTTGGCAAGAAAGAGAGGTG 3300
4017 AAAATAGAAAAATAATGCCATGTTATTCGGGCTGTGTCAGAAATCTACTGCTTTAACC 4076
3301 AAAATAGAAAAATAATGCCATGTTATTCGGGCTGTGTCAGAAATCTACTGCTTTAACC 3360
4077 CAGCACAGGGGAAAAACCATAGAACTCTCAGAGAGCCCTAGACCCACTTTTCATGGAT 4136
3361 CAGCACAGGGGAAAAACCATAGAACTCTCAGAGAGCCCTAGACCCACTTTTCATGGAT 3420
4137 CCAGACTTGGCATATGGAACTTATACAGGAAGCTGTGGTCATGTAATGACGAGGTGTC 4196
3421 CCAGACTTGGCATATGGAACTTATACAGGAAGCTGTGGTCATGTAATGACGAGGTGTC 3480
4197 TGGCAGAAAGTATTTTGAAGCTGTACAGCTGAGCTCTCAGCAGCGCATTCATGTTGACCTT 4256

Db	3481	TGCAGAAATATTTTGAAGCTGTACAGCTGAGCTCTCAGAGCGCAATTCATGTTGACCTT	3540
Qy	4257	TTTGACATTGGAAGTGGAGAAATATCTTTGGCCCTCTTGCAAAATCTCTGTGCAATACTGTG	4316
Db	3541	TTTGACATTGGAAGTGGAGAAATATCTTTGGCCCTCTTGCAAAATCTCTGTGCAATACTGTG	3600
Qy	4317	ATCCCCATATTCCCTTTGCAACCTCAAAAGATAACAGTGAAGATGCAGATGCTCTTGCT	4376
Db	3601	ATCCCCATATTCCCTTTGCAACCTCAAAAGATAACAGTGAAGATGCAGATGCTCTTGCT	3660
Qy	4377	CAACTTTTGACCTTGGCACCGTGGATACAGACTCTTCTGGCCAGAAATATCAGGTTATAAT	4436
Db	3661	CAACTTTTGACCTTGGCACCGTGGATACAGACTCTTCTGGCCAGAAATATCAGGTTATAAT	3720
Qy	4437	ATAAGACATGCTAAAGGAGAAAACCCCAATTCCTATTTCCTTTAATCAAGGAATGGAGAT	4496
Db	3721	ATAAGACATGCTAAAGGAGAAAACCCCAATTCCTATTTCCTTTAATCAAGGAATGGAGAT	3780
Qy	4497	TCTACTTTGGAGTTCCATTCATCCTGAGTTTGGCGTTGAGTCTTCGATTTAAATATCA	4556
Db	3781	TCTACTTTGGAGTTCCATTCATCCTGAGTTTGGCGTTGAGTCTTCGATTTAAATATCA	3840
Qy	4557	AATAGCATCAAGGAAATGGTTATTCTCTTGGCCACAACAATTTATAGAAATGGATTGAAA	4616
Db	3841	AATAGCATCAAGGAAATGGTTATTCTCTTGGCCACAACAATTTATAGAAATGGATTGAAA	3900
Qy	4617	GTGCCACTGATGAAGGAGATCTCGAGTCCCCATGTGACCTGGAGCACTCGCGCTTTC	4676
Db	3901	GTGCCACTGATGAAGGAGATCTCGAGTCCCCATGTGACCTGGAGCACTCGCGCTTTC	3960
Qy	4677	ACTATCAGGCAATTGAAATCTATTGGGAGATCAAGGAAACCTCTGTTTGGAGCACTT	4736
Db	3961	ACTATCAGGCAATTGAAATCTATTGGGAGATCAAGGAAACCTCTGTTTGGAGCACTT	4020
Qy	4737	CAAAATAGGCAATATGCTGCTGAAAGCATTAATGCAGTTTGCAGTTGCACAGAGGAT	4796
Db	4021	CAAAATAGGCAATATGCTGCTGAAAGCATTAATGCAGTTTGCAGTTGCACAGAGGAT	4080
Qy	4797	ACCTGTCTCAGGTCCTGATACAGAAACATCTGTTGCTCTCTCATCATGTTGTTCTTCT	4856
Db	4081	ACCTGTCTCAGGTCCTGATACAGAAACATCTGTTGCTCTCTCATCATGTTGTTCTTCT	4140
Qy	4857	AACATAAATCAGAAATACACCATGCTTCTGCTCTATAGATCTGTTTCATGTTTGGTG	4916
Db	4141	AACATAAATCAGAAATACACCATGCTTCTGCTCTATAGATCTGTTTCATGTTTGGTG	4200
Qy	4917	GGTGTGTGTAGCATTTCCCATCTCTGTTGGGATGACCTGTTGATCTGCAGCTTCT	4976
Db	4201	GGTGTGTGTAGCATTTCCCATCTCTGTTGGGATGACCTGTTGATCTGCAGCTTCT	4260
Qy	4977	TCAGTTAGTTCTTCTTATACCACTTTATCTCTTCCATTTGATCACCATGGCACACATG	5036
Db	4261	TCAGTTAGTTCTTCTTATACCACTTTATCTCTTCCATTTGATCACCATGGCACACATG	4320
Qy	5037	CTTCAGATACACTTACAGTACACAGGCTTACCCCTTGTCTGCTAGATCTGTTTCAGAGACAGT	5096
Db	4321	CTTCAGATACACTTACAGTACACAGGCTTACCCCTTGTCTGCTAGATCTGTTTCAGAGACAGT	4380
Qy	5097	GAAGGGCTCATTCGGCATCTTCTTCTTCGAGAAATTTCTCAATATACAGTGGCTTCC	5156
Db	4381	GAAGGGCTCATTCGGCATCTTCTTCTTCGAGAAATTTCTCAATATACAGTGGCTTCC	4440
Qy	5157	ATTGGGTGTGATATTCTCGGCTGTGTTTGTGGGTCTCACATGAAGATGGCATCACCT	5216
Db	4441	ATTGGGTGTGATATTCTCGGCTGTGTTTGTGGGTCTCACATGAAGATGGCATCACCT	4500
Qy	5217	TATCTTCGCTGTGCTGCTTTTTCACATATTTTACCTTGGGTTAACTCCGCTGAGGAA	5276
Db	4501	TATCTTCGCTGTGCTGCTTTTTCACATATTTTACCTTGGGTTAACTCCGCTGAGGAA	4560
Qy	5277	CTGCATACCAATTTCTGCAGAGGAGAGTACAGTGCAGTCTGTAGCTATCTATCTTTACCT	5336
Db	4561	CTGCATACCAATTTCTGCAGAGGAGAGTACAGTGCAGTCTGTAGCTATCTATCTTTACCT	4620

RESULT 2
AAV99308
ID AAV99308 standard; cdNA; 6395 bp.

Qy	5337	ACAAATTTTGTCTCTCTCTTCCAGGAATATTTCGGGATACTGTAAAGGCCCTTGTCTCCAGAGG	5396
Db	4621	ACAAATTTTGTCTCTCTCTTCCAGGAATATTTCGGGATACTGTAAAGGCCCTTGTCTCCAGAGG	4680
Qy	5397	TGGTGTGCAGATCCCTTCTTACTAACTGTTTGAAGCAAAACACACCTGGTGCAGGTAC	5456
Db	4681	TGGTGTGCAGATCCCTTCTTACTAACTGTTTGAAGCAAAACACACCTGGTGCAGGTAC	4740
Qy	5457	CCTAGAAAAGAAATAGTTTGTATAGAGCTTCTCTGATGACTATAGCTGCTCTCTGAATCAA	5516
Db	4741	CCTAGAAAAGAAATAGTTTGTATAGAGCTTCTCTGATGACTATAGCTGCTCTCTGAATCAA	4800
Qy	5517	GCTTCTCATTTTCAGGTGCCACGGTCTGCAGATGATGAGCGAAGCATCCTGTCTCTGTC	5576
Db	4801	GCTTCTCATTTTCAGGTGCCACGGTCTGCAGATGATGAGCGAAGCATCCTGTCTCTGTC	4860
Qy	5577	CTTTTCTGTGGGCTTACTATGTTCTCAGAACATTTGCTGCCAGAAATTTGTAACGGG	5636
Db	4861	CTTTTCTGTGGGCTTACTATGTTCTCAGAACATTTGCTGCCAGAAATTTGTAACGGG	4920
Qy	5637	GAAGAGTTGGAGCTTGCATTTTTCACGCACCTTCACTGTGGAGCCGAGTCTGCAATTTTC	5696
Db	4921	GAAGAGTTGGAGCTTGCATTTTTCACGCACCTTCACTGTGGAGCCGAGTCTGCAATTTTC	4980
Qy	5697	CTAAAATCAGAGAATGCCGAGTGGTCTCTGTTGAAGGTAAAGCCAGAGGCTGTGCCTAT	5756
Db	4981	CTAAAATCAGAGAATGCCGAGTGGTCTCTGTTGAAGGTAAAGCCAGAGGCTGTGCCTAT	5040
Qy	5757	CCAGCTCTTACTTCTGGATGAATATGAGAAACAGACCCCTGGCTGAAGAGGGCAACCCC	5816
Db	5041	CCAGCTCTTACTTCTGGATGAATATGAGAAACAGACCCCTGGCTGAAGAGGGCAACCCC	5100
Qy	5817	CTTCAATTTATCTGAGCGGTATCGGAAGCTCCATTTGCTGCGCAACAACACTGCATTT	5876
Db	5101	CTTCAATTTATCTGAGCGGTATCGGAAGCTCCATTTGCTGCGCAACAACACTGCATTT	5160
Qy	5877	ATAGAAGATTTGCTAGGAGCCAAAGACTAATACAGATGTTATTTGGATTCAACTGGCAG	5936
Db	5161	ATAGAAGATTTGCTAGGAGCCAAAGACTAATACAGATGTTATTTGGATTCAACTGGCAG	5220
Qy	5937	TTACTGTGAGCTTCCAACTCTGCTCAAGACAATCACAATGACGACAGTAGTAAAGGCTG	5996
Db	5221	TTACTGTGAGCTTCCAACTCTGCTCAAGACAATCACAATGACGACAGTAGTAAAGGCTG	5280
Qy	5997	ATTCAAAATTTATGGAACCTTTCTGAGGCTGGGAAAGTATTGGAGGGTCTTTTGTCTCCA	6056
Db	5281	ATTCAAAATTTATGGAACCTTTCTGAGGCTGGGAAAGTATTGGAGGGTCTTTTGTCTCCA	5340
Qy	6057	TGTCAGGTTCACTTACATCAATAAATAATTTCTTAATGGAGTATTGCTTTCAATTAGCA	6116
Db	5341	TGTCAGGTTCACTTACATCAATAAATAATTTCTTAATGGAGTATTGCTTTCAATTAGCA	5400
Qy	6117	ACATATGCTTACAGGAAAAAGGACATAGATCAATCTGTTTATGCTAGTATTTC	6176
Db	5401	ACATATGCTTACAGGAAAAAGGACATAGATCAATCTGTTTATGCTAGTATTTC	5460
Qy	6177	AGGAATTTATCCCTTTCATTAATTTCTCATTTTATTTTTCATCCACTGGTAGAT	6236
Db	5461	AGGAATTTATCCCTTTCATTAATTTCTCATTTTATTTTTCATCCACTGGTAGAT	5520
Qy	6237	GAAGTCACTCAACACAGTTTGTAGACATTTTATGTTGGTTAACTCTTCTGCAATTTTGT	6296
Db	5521	GAAGTCACTCAACACAGTTTGTAGACATTTTATGTTGGTTAACTCTTCTGCAATTTTGT	5580
Qy	6297	ATTTGGTGTGTTT 6308	
Db	5581	ATTTGGTGTGTTT 5592	

Qy	2029	TGCTAGAAGTTTACCTGAGTACTTGGACAGGAACAATAAATTCAACTTCCAGGTTATA	2088	
Db	1448	TGCTAGAAGTTTACCTGATACTTGGCAGGACAAATAATTCAACTTCCAGGTTATA	1507	
Qy	2089	GCCAGGACAAATGGGAAGAGTATATGACAGTAATATGACCTAAAGTATATCTGATCA	2148	
Db	1508	GCCAGGACAAATGGGAAGAGTATATGACAGTATATGACCTAAAGTATATCTGATTA	1567	
Qy	2149	GCAAAACCCACAATATGACAGAAAGATTAAGAATGAGTTCCTTGAAGGTTTTCGATCTT	2208	
Db	1568	GCAAGCCTGTATATGACAGAAAGATTAAGAAGCGCATGTCCTGGAAGGTTTCGGGCTT	1627	
Qy	2209	TTTTGAAGATTCTTACCTGTATGACGGAATGGGAAGAAATCCGAAGACAGTTGGGCAAC	2268	
Db	1628	TTCTGAAGATTCTTACCTGTATGACGGAATGGGAAGAAATCAGAGCAAGTTGGACAAC	1687	
Qy	2269	ACATTGAAGTGGATTCCTGATTGGGAGGCTGCATTTGATACAGATGCAATTTGAAGAATA	2328	
Db	1688	ACATTGAAGTGGACCTGACTGGGAGGCTGCCATCGCTATACAGATGCAACTAAGAATA	1747	
Qy	2329	TTTTTACTCATGTTCCAAAGAGTGGTGTGTGTGATGAAGAACTCTTACTTGTGGCTTATA	2388	
Db	1748	TTTTTGTCTCATGTTCCAAAGAGTGGTGTGTGTGTGATGAAGATCTCTTACTTGTGGCTTATA	1807	
Qy	2389	AAGAACTGTCAAAAGCTGTGATGAGGTGCAGTACCAGTTTCATATCTAGTAGCAAGACAG	2448	
Db	1808	AAGAACTGTCAAAAGCTGTGATGAGGTGCAGTACCAGTTTCATCTCCAGTACCAGACAG	1867	
Qy	2449	TAGTACAAATCGTGTGGACATAGTTTGGAAACAAAGTCTCAGAGAGTATCTCAGGATCTTG	2508	
Db	1868	TAGTGCAAATGTGGGTCATAGTCTGGAACCAAAATCCTACAAAGTGTCTGAGGACCTTG	1927	
Qy	2509	TAAGCATACATCTGCCACTCTCTAGGACCCCTTGCTGTCTTCATGTAGCTTTAAGCAGGC	2568	
Db	1928	TAAGCATACACCTGCCACTCTCTAGAACACTTGCTGTCTTCATGTAGCTTTAAGCAGAC	1987	
Qy	2569	TGGTGTCTGTTTCAAGACTGATGAATTTGTGTCTTTTGAGGACTTTCAAGTAGAGTAC	2628	
Db	1988	TAGTGTCTATTTCAAGACTGATGAATTTGTGTCTTTTGAGCCTTTTCAAGTAGAGTAC	2047	
Qy	2629	TAGTGGAAATATCCTTTACGTTGTCTGTGTGTGGTGGCCAGGTTGTGCTGAGATGTGGC	2688	
Db	2048	TGGTGGAGTACCCGCTGCGTGTGTGTGTGGTGGCTCAGGTTGTGCTGAGATGTGGC	2107	
Qy	2689	GAAGAAATGGACTGTCTTTATAGCCAGGTGTTTATACCAAGATGTTTAAAGTCAGAG	2748	
Db	2108	GAAGAAACGGCTCTCACTCATCAGCCAGGTTTCTATTATCAAGATGTTTAAAGTCAGG	2167	
Qy	2749	AAGAAATGTATGATAAAGATATCATCATGCTTCAGATTTGTTGATCTTTAATGATCCCA	2808	
Db	2168	AGGAAATGTACGATAAAGATATCATCATGCTTCAGATTTGAGAGTCTTATATGATCCCA	2227	
Qy	2809	ATAAGTTCTTGTACTTGTACTTACAGAGTATGAACTTGGCAGGCTTTTAAAGAACCA	2868	
Db	2228	ACAAGTTCTTGTACTTGTACTTACAGAGATGAACTTACTGTGCTTTTAAAGAACCA	2287	
Qy	2869	TATCTACAAAGACAGGATTTGATTAACAAATATAATACACTAATAGAAGAAATGCTTC	2928	
Db	2288	TATCCAAAGACAGGATTTGATTAACAAATATAATACACTAATAGAAGAAATGCTTC	2347	
Qy	2929	AGGTCTCTCATCTATATTGTGGGTGAGGCTTATGTACCTGGAGTGGGAAATGTGACCAAG	2988	
Db	2348	AGGTCTCTCATCTATATTGTGGGACAGCTTATGTACCTGGAGTGGGAAATGTTACCAG	2407	
Qy	2989	AAGAGGTACAAATGAGAAATCATCTTGTCTTTCATTTGAACCCATGCCACACAGTG	3048	
Db	2408	AGGAGGTTAATAATCAGAGAGATTACTCACTTTACTTTTGCATTTGACCCATGCCACAGTG	2467	
Qy	3049	CCATTGCCAAAATTTACCTGAGAAATGAAAATATGAAACTTGGCTTAGAGAAATGTCATA	3108	
Db	2468	CCATCGCAGAAACCTACCTGAGAACGAAAATATGAAACTTGGCTTAGAGAAATGTCATA	2527	
Qy	3109	ACAAAGTGGCCACATTTAAGAAACACAGGTGTATCAGGCCATGGAGTTTATGAACTAAAG	3168	
Qy	2528	ACAAAGTGGCCACATTTAAGAAACACAGGTGTGTCGGGCCATGGAGTTTATGAAATGAAAG	2587	
Qy	3169	ATGAATCACTGAAAGACTTCAATATGTATTTTATCATTTACTTCTCAAAACCCAGCATPAGA	3228	
Db	2588	ATGAATCACTGAAAGACTTCAATATGTATTTTATCATTTACTTCTTAAACACACAGCATPAGA	2647	
Qy	3229	AGGCTGAACATATGACAGAAAGAGAGAAACAAAGAAAGATGAAGCATTTGCCGC	3288	
Db	2648	AGGCTGAACATATGACAGAAAGAGAGAAACAAAGAAATGAAGATGAAGCATTTGCCGC	2707	
Qy	3289	CACCAACACCTCTCAATTCGCTCTTTCAGCAAAAGTGATTAACCTTCTCAACTGTG	3348	
Db	2708	CGCCACCTCTCCAGAGTTCTGCCCTGCTTTCAGCAAAAGTAGTCAACCTGCTCAGCTGTG	2767	
Qy	3349	ATATCATGATGTACATTTCTCAGGACCGTATTTTGGCGGCAATAGACAGCATTTTAACT	3408	
Db	2768	ATGTTATGATATACATCTCAGGACCATCTTTGAGCGGCGAGTGGACACGGAGTCTTAATC	2827	
Qy	3409	TGTGACCGAAGGGATGCTCCAAATGGCTTTTCATATTTCTGCAATTTGGCTTTTACTAGAAG	3468	
Db	2828	TGTGACAGAGGGATGCTGCAAGTGGCTTCCATATATTTGGCACTTGGCTTGTGGAAG	2887	
Qy	3469	AGAAGCAACAGCTTCAAAAGCTCTGAGAGAAAGTAACTTTGACTTTTATCATAGG	3528	
Db	2888	AGAAGCAACAGCTTCAAAAGCTCTGAGAGAAAGTGGCTTTGACTTTTACCATAAAG	2947	
Qy	3529	CTTCAAGATTGGGAAGTTTCAAGCAT-----GAATATACAAATGCTTTTGGGAAAC	3579	
Db	2948	CTTCAAGATTGGGAAGTTTCAAGCATGATGCTCAGAAATATACAAATGCTTTGGGAAAGAC	3007	
Qy	3580	TCAAAGGAATTTCCCGAGTTAGAGCCAGAGGACATGATTAACGTGGATCTTCAAGATGT	3639	
Db	3008	TCAAAGGAATTTCCCGAATTTAGAGCCAGAGGACATGATTAACGTGGATCTTCAAGATGT	3067	
Qy	3640	TTGACACAGTGAAGCGATTTAGAGAAATCTTTGTTTAAATTTAGCAACCATCATAGGAT	3699	
Db	3068	TTGACACAGTGAAGCGATTTAGAGAAATCTTTGTTTAAATTTAGCAACCATCTTCAAGAC	3127	
Qy	3700	CGGAATCTATTAAAGATGATGATTAACATGATTAACAAAGAAACAGAACCAAGAA	3759	
Db	3128	TGGAGTGAATTAAGAGTGAAGGATTTACTTATGATAAAGAAAGGAGGAGGAGAGAA	3187	
Qy	3760	AAGCTGAAGCTGCTAGGCTACATCGCCAGAGATCATGCTCAGATGTCTGCTTACAGA	3819	
Db	3188	AAGCTGAGCCCTAGGCTTTCATGCCAGAGATCATGSCCCAGATGTCTGCTTACAGA	3247	
Qy	3820	AAACTTCAATGAAACTCATAACTCATGTATGACAAATACATCAAGAAATGCTTGGGAAAG	3879	
Db	3248	AAACTTCAATGAAACCCCAAACTCATGTATGATAAATACGTGAGAAATGAGGAGG	3307	
Qy	3880	AGATTTCCATTTAGGAGGAGAGACACCCAGCAGTCACTGACTACTTAGAATTTGCTT	3939	
Db	3308	AAGCTTCCATTTAGGAGGAGAGACACCTCAGCAGTCACTGAGGCTCTAGAATTTGCTC	3367	
Qy	3940	TGGTCTTAAACGGGCTCCATCTGTTTACTGAAAAGGAGGTGCTGAGCTGCTCTTGGC	3999	
Db	3368	TGGGCCCTTAAACGGGCTCCGCTGTTTACCAGAAAGAGGAGTGTGAGCTGCTCTGCTG	3427	
Qy	4000	AAGAAAGACAGGAGGTGAAATAGAAATTAATGCCATGTTATTCGGCTCTGTGTCAGA	4059	
Db	3428	AAGAAAGACAGGAGGTGAAATAGAAATTAATGCCATGTTATTCAGCATGTGTGTCAGA	3487	
Qy	4060	AATCTACTGCTTAAACCCAGCAGGAGGAAACCCATAGAACTCTCAGGAGAACCCCTAG	4119	
Db	3488	AATCCACCCCTTAAACCCAGCAGGAGGAGGCTGTGGACCACTTAGGGGAAACACTGG	3547	
Qy	4120	ACCACTTTTTCATGATCCAGACTTGGCATATGGAATTTATACAGAACTGTGCTCATG	4179	
Db	3548	ACCTCTTTTTCATGATCCAGACTTGGCATATGGAATTTATACAGAACTGTGCTCATG	3607	
Qy	4180	TAATCAGCGAGTGTGCTGGCAGAGATTTTGAAGCTGTACAGCTGTGCTCTCAGCAGC	4239	

Db 3608 TAATGCATGCAGTGTCTGGCAGAGTATTTTGAAGCTGTGCAGCTGAGCTCGCAGCAGC 3667
Qy 4240 GCATTCATGTTGACCTTTTGGACTTGGAAAGTGGAGAAATATCTTTGCCCTCTTTGGCAAT 4299
Db 3668 GCATTCACGTAGACCTGTTTGAACCTGGAGAGCGCGAGTAGTACCTATGCGCCCTCTGCAAGT 3727
Qy 4300 CTCTGTGCAATAGTGTGATGCCCATATTCCTTTTGAACCTCAAAAGATAAAGAGTGTGAG 4359
Db 3728 CTCTGTGCAACACTGTCAATCCCATCATCTTTTGCAGCGCAGAGATCAACAGTGTAGA 3787
Qy 4360 ATGCAGATCTCTTGTCTCAACTTTTGGACCTGTGCAGCGGTGGATACAGACTGTTCTGCGCCA 4419
Db 3788 ATGCGAGGCTCTTGTCTCAACTTTTGAACCTGTGCAGCGGTGGATACAGACTGTTCTGCGCCA 3847
Qy 4420 GAATATCAGGTATATAATAAGACATGCTAAAGGAGAAACCC---CAATTCCTATTTTCT 4476
Db 3848 GAATATCGGGTTATAATAAAGCATGCTAAAGGAGAAAGCCCGCAGAGTTTCTGTCTGT 3907
Qy 4477 TTAATCAAGGAATGGAGATTTCTACTTTTGGAGTTTCCATTCCTGAGTTTGGCGTTG 4536
Db 3908 TTAATCAAGGAATGGGGATTTCAACTTTTGAAGTTTCAATTCCTGAGTTTGGAGTTC 3967
Qy 4537 AGTCTTCGATTAATATTCAAATAGCATCAAGGAAATGTTATTTCTTTTGCACACAA 4596
Db 3968 AGTCTTCGTTGAATATTCAAATAGTATCAAGGAAATGTTATTTCTTTGCGCCACAA 4027
Qy 4597 TTTATAGAAATGGATTTGAAGTGGCCACCTGATGAAAGGGATCTCGAGTCCCATGCTGA 4656
Db 4028 TTTACAGAAATGGCCGTGAAAGTGCCTCTGATGAACCTAGACCCACGAGTGCCTATGATGA 4087
Qy 4657 CTTGAGACACCTGCGGCTTTCATATCCAGGCAATTTGAAATCATTTTGGGAGATGAAGGAA 4716
Db 4088 CTTGAGACACCTGCGGCTTTCATATCCAGGCAATTTGAAATCATTTTGGGAGATGAAGGAA 4147
Qy 4717 AACCTCTGTTTGAGACATTTCAAAATAGCAGCATATGCTGTGAAGCATTTAATCCAGT 4776
Db 4148 AACCTCTATTTGGAGCATTCAAAATAGACAGCATAGCGGCTGTGAAGGCGCTAATGCAGT 4207
Qy 4777 TTGCAAGTTCACAGAGGATTTACCTGTCTCAGGCTCTGTATACAGAAACATCTGGTCTGTC 4836
Db 4208 TTGCAAGTTCACAGAGGCTTACCTGTCTCAGGCTCTGTATACAGAAACATCTGGTCTGTC 4267
Qy 4837 TTCTATCAGTTGTTCTTCTTAACATATAAATAGAGATACACCATGCTCTGCTCTATAG 4896
Db 4268 TCTCTGCACTTATCTTCTTAACATCAGAAATACACAGGCTTCTGCTGCTGG 4327
Qy 4897 ATCTGTTTCATGTTTGGTGGGTGCTGTGTAGCATTCCTATCCATGTTATGGGATGACC 4956
Db 4328 ATCTCTTCCATGTTCTGTTGGTGGGCGAGTCTTAGCGTTCCCATCTGTTATTTGGGATGACA 4387
Qy 4957 CTGTTGATCTGCAGCCTTCTTCAAGTTAGTCTTCTTCTTATACACACCTTTATCTCTCCATT 5016
Db 4388 CCGTGATCTGCAGCGCTGCGCACTTAGTCTTCTATATACACACCTTCTATCTCTCCATC 4447
Qy 5017 TGATACCATGGCACATGCTTCAGATCTACTTACAGTAGACA-----CAG 5064
Db 4448 TGATACCATGGCGCACATGCTTCAGATCTACTTACAGTAGACA-----CAG 4507
Qy 5065 GCTACCCCTTGTCTCAGGTTCAAGAGACAGTGAAGAGGCTCAATCCGCACTTCTTCTTCT 5124
Db 4508 GCGCGCGCTTGTCTGAGGTTGAAGAGATAGTGAAGAGGCTCCTGTGCACTGCTTCT 4567
Qy 5125 TTGCAAGAAATTTCTCAATATACAAAGTGGCTTCATTTGGGTGTGATATTCCTGGGTGTTAT 5184
Db 4568 TTGTGGAAGTGTGCGACACAGAGCGGCTCACTGGGTGCGGTGCTCCCGGCTGGTACC 4627
Qy 5185 TGTGGGTCTCACTGAAGATGGCATCACCCCTTATCTTCTGCTGTGCTGATTTTCTTCC 5244
Db 4628 TGTGGGTCTCCCTGTGAGAGCGGCATCACCCCTTACCTCCGCTGTGCTGCACTCTTCC 4687
Qy 5245 ACTATTTACTTGGGTGAACCTCCGCTTGAGAACTGCATACCAATTTCTGCAAGAGGAGAT 5304
Db 4688 ACTATTTACTTGGAGTAGCTCCGCTTGAAGAACTGTTTGGCAATTTCTGCTGAAGAGAGAT 4747

Qy 5305 ACAGTGCACCTCTGTAGCTATCTATCTTTACCTACAAAATTTGTTCTGCTCTTCCAGGAAT 5364
Db 4748 TCAGTGCACCTCTGTAGCTATCTATCTTTACCCACAAAATTTGTTCTGCTCTTCCAGGAAT 4807
Qy 5365 ATTGGGATCTGTAAAGCCCTTGTCTCCAGAGTGTGTGCAGATCTGCTTCTACTTAACT 5424
Db 4808 ATTGGGATCTGTAAAGCCCTTGTCTCCAGAGTGTGTGCAGATCTGCTTCTACTTCAAGT 4867
Qy 5425 GTTTGAAGCAAAAACACCCTGCTAGGTACCTTAGAAAAGAAAATAGTTTGTATAGAGC 5484
Db 4868 CTTTGAAGCAAAAAGTCTGTGTGTCAGGTACCTTAGAAAAGAAAATAGTTTGTATAGAGC 4927
Qy 5485 TTCTGTAGTACTATGATGCTCTCTGAAATCAAGCTTCTCAATTCAGGTGCCACCGGTCTG 5544
Db 4928 TTCTGTAGGACTACAGCTGTCTTCTAAATCAGGCTTCTCACTTTAGGTGTCCACGCTCTG 4987
Qy 5545 CAGATGATGAGCAAGAGCATCTCTCTCTGCTCTGCTTTTCTGTGGGCTATACTATGTTCTC 5604
Db 4988 CAGATGATGAGCAAGAGCATCTCTCTCTCTGCTTTTCTGTGGGCTATCTGTTCTCTC 5047
Qy 5605 AGAATCTTGTCTGCCAGGAAATTTGTAAAGGGGAGAGTGTGGAGCTTGCATTTTTCAGC 5664
Db 5048 AGAATCTTGTCTGCCAGGAAATTTGTAAAGGGGAGAGTGTGGAGCTTGCATTTTTCAGC 5107
Qy 5665 CACTTCACTGTGGAGCGGAGTCTGCAATTTTCTTAAATAATCAGAAATGCCAGTGGTCC 5724
Db 5108 CGCTTCACTTGTGGTCTGAGCTCTGCAATTTTCTTAAATAATCCGAGAATGCAGGTGGTCC 5167
Qy 5725 TGGTTGAAGTAAAGCCAGAGGCTGTGCTATCCAGCTCTTACTTGTGATGAATATGAG 5784
Db 5168 TGGTTGAAGGAAAAGCCAGAGGCTGTGCTTACCAGCCCTTACTTGGATGAATATGAG 5227
Qy 5785 AAACAGACCTTGGCTTGAAGAGGGCAACCCCTTCAATTTATCTCGTGAGCGGTATCGGA 5844
Db 5228 AAACAGACCCAGGCTTAAAGAGAGAAACCCATCTTCAATTTATCTCGGAGCGGTATCGGA 5287
Qy 5845 AGCTTCCATTTGGTCTGGCAACACACTGCAATTTATAGAAGAGATTTCTTAGGCCAAGAGA 5904
Db 5288 AGCTGCAATTTGGTCTGGCAACACACTGCAATTTATAGAAGAGATTTCTCGGAGCCAGAGA 5347
Qy 5905 CTATCAGATGTTATTTGGATTTCAACTGGCAGTACTGTGAGCTCCAATCTGCTCTCAAG 5964
Db 5348 CTATCAGATGCTATTTGGATTTAACTGGCAGTACTGTGAGCTTCACTTCTGCTCTCAAG 5407
Qy 5965 ACATCAAAATGACGACAGTAGTAAAGGCTGATTCAAAATATGGAATACTTCTGAGG 6024
Db 5408 ACAATCATGATGACATCAATAAAGAGCTGATCTAAAATTTCTAGAGAACTTTCTGAGG 5467
Qy 6025 GCTGGGAAAGTATTTGGAGGCTTTTGTCTCCATGTCAGGTTCACTTACATCAATAAAT 6084
Db 5468 ACGGGGAAGTATTTGGAGGCTTTTGTATCCATGTCAGAGTTCAACACATTAATAAAT 5527
Qy 6085 ATTTCTTAATGGAGTATTTGCTTCAATTTAGCAACATATGCTTCAAGGAAAAA-AGGAC 6143
Db 5528 ATTTCTTAATGGATATTTGCTTCAATTTATCAACATAGCTTCAAGGAAAAAACAAGAC 5587
Qy 6144 ATAGATCAATCTGT-----TTT 6160
Db 5588 ATAGATTAATGTTTTATGTTCTAGAACACTAAAGAAATGCTTGTTCATCCAAAGTCTCTAT 5647
Qy 6161 ATGCTGATGATTTTCCAGGAATTTATTTCCCTTCAATAATTTGCTCTCATTTTCAATTTAT 6220
Db 5648 TTCTGCTTAATTTTCCAGAAAACTCTTTCCCTTCAATAAGTGTCTAGTTTCAATTTCAAT 5707
Qy 6221 CATCCATGCTAGATGAAGTCAAGTCAAGAGTGTAGACATTTTATGTTGTTGTTAAC 6280
Db 5708 CACCCACTGTTTAAATGAGGTCACATTAAGCATTTTGTGGACATTTCTCCATCTGGCTAAC 5767
Qy 6281 TCTTCTGCAATTTTGTATTTGGTGT 6307
Db 5768 ATCTCTGCACCTTTGTATTTGGTGT 5794

```

RESULT 3
AAC86933 standard; cDNA; 6395 BP.
XX AC AAC86933;
XX DT 02-APR-2001 (first entry)
XX DE Nucleotide sequence of Mouse Ub1 protein.
XX KW Ub1; E3-type protein; ubiquitin system; ubiquitin-protein ligase;
XX KW N-end rule pathway; intracellular pathogen; Yersinia monocytogenes;
XX KW Yersinia enterocolitica; muscle wasting; infection; ss.
XX OS Mus sp.
XX FH Key Location/Qualifiers
XX FT 115..5388
XX FT /*tag= a
XX FT /product= "Ub1"
XX PN US6159732-A.
XX PD 12-DEC-2000.
XX PP 11-JAN-1999; 99US-0228317.
XX PR 02-DEC-1997; 97US-0982956.
XX PA (CALY ) CALIFORNIA INST OF TECHNOLOGY.
XX PI Kwon YT, Varshavsky A;
XX FI WPI; 2001-090278/10.
XX DR P-PSDB; AAB31162.
XX PT Inhibiting the N-end rule pathway in mammalian cells for treating
XX PT infections and various diseases associated with muscle tissue wasting,
XX PT by inhibiting the expression of Ub1 gene
XX PS Claim 3; Column 15-28; 18pp; English.
XX CC The present sequence encodes a murine Ub1 enzyme. Ub1 is an E3-type
XX CC protein of the ubiquitin system. Specifically, it is a ubiquitin-protein
XX CC ligase. The enzyme is specific for destabilising residues exposed at
XX CC the N-terminus of protein substrates. Inhibition of the expression of
XX CC Ub1 gene in a cell results in inhibition of the N-end rule pathway.
XX CC The method is used for treatment of mammalian cells infected with an
XX CC intracellular pathogen, e.g. Yersinia monocytogenes or Yersinia
XX CC enterocolitica. Inhibition of N-end rule pathway is also useful for
XX CC treating various diseases associated with wasting of muscle tissue and
XX CC infections.
XX SQ Sequence 6395 BP; 1802 A; 1376 C; 1519 G; 1698 T; 0 other;

Query Match 70.3%; Score 4435.4; DB 22; Length 6395;
Best Local Similarity 86.3%; Pred. No. 0;
Matches 4997; Conservative 0; Mismatches 721; Indels 69; Gaps 6;

QY 590 TTCAGGGCCCTCGTAAAGTGTGTCCTCTGTC-TCCGACCGGCCACAGTTTCCTCGCT 648
DB 8 TTCAGGGCCCTCGTAAAGTGTGTCCTCTGTC-TCCGACCGGCCACAGTTTCCTCGCT 67
QY 649 TGCCTCTGGCCGGGGTGGGCACTGCAGGCGTCAAGTTTCCTCAAGATGCGGGACGAGG 708
DB 68 AGCTGGCGGGCGGGGTGGGAACTGGCGGGTTCGTTCCCTTAAGATGGCGGACGAAG 127
QY 709 AGGCTGGAGGTACTGAGAGGTGGAATACAGCGGGAGTTACCCACAGACCCCTCAGCGTC 768
DB 128 AGATGGAGGGCGCGGAGAGGTGACGTCAGCCCGGAGCCCTCCCTCGCCCGGACGCGC 187
QY 769 TGGCATCTTGTGGGATCAGCAAGTTGATTTTATCTACTGCTTCTTCATCATTTGGCAC 828

```

Db	1268	AAAACTCTTTGCTATGGAATTTGTGAAGTATTATTAACAACATGCGACAAGAGTACATCA	1327
Qy	1909	GTGATGATCATGACAGAACTATCTCTATAAATGCACTTTTCAGTTCAGATGTTTACGTGTC	1968
Db	1328	GGGACGACCACGAGAGAACATCTCCATAAACGCCCTGTCCGTCAGATGCTCACCGTCC	1387
Qy	1969	CTACTCTGGCTCGACATCTTATTAAGACGACAAATGTTATCTCTGTCATTTACTGAAACTC	2028
Db	1388	CGACCTTGGCCCGGACTCTTATTTGAAGACGACAAATGTTATTTCTGTCATTTACTGAAACGC	1447
Qy	2029	TGCTAGAAGTTTAACTCTGAGTACTTGGACAGAACAAATAAATTCAACTTCCAGGGTTATA	2088
Db	1448	TGCTAGAAGTTTAACTCTGAGTACTTGGACAGAACAAATAAATTCAACTTCCAGGGTTATA	1507
Qy	2089	GCCAGGACAAATTTGGGAAGATATATGCAATATATGACCTTAAAGTATATCTCTGATCA	2148
Db	1508	GCCAGGACAACTGGGAAGTCTAGCGAGTTATATGTGACCTTAAAGTATATCTCTGATTA	1567
Qy	2149	GCAAAACCCACAATATGGACAGAAAGATTAAGAAATGCAAGTTTCCAGGGTTTTCATCTT	2208
Db	1568	GCAAGCCTGTCTATATGGACAGAACGATTAAGAGCGCAGTTCTCTGGAAGGTTTCCGGTCTT	1627
Qy	2209	TTTTGAAGATCTTACCTCTGATGCAAGGAATGGAAGAAATCCGAACAGGTTGGCAAC	2268
Db	1628	TTCTGAAGATCTTACCTGTATGCAAGGAATGGAAGAAATCGAAGACAAAGTTGGCAAC	1687
Qy	2269	ACATTTGAAGTGGATCTGATTTGGGAGCTGCCATTTGCTATACAGATGCAATTCGAAGATA	2328
Db	1688	ACATTTGAAGTGGACCTGACTGGAGGCTGCCATCGCTATACAGATGCAATTCGAAGATA	1747
Qy	2329	TTTTTACTCATGTTCCAAAGAGTGGTGTGTTGATGAAGAACTCTTACTTTGGTGGCTTATA	2388
Db	1748	TTTTTGTCTATGTTCCAAAGAGTGGTGTGTTGATGAAGAACTCTTACTTTGGTGGCTTATA	1807
Qy	2389	AGAATGTCACAAAGCTGTGATGAGGTGCAGTACCAAGTTTCATATCTAGTAGCAACAG	2448
Db	1808	AGAATGTCACAAAGCTGTGATGAGGTGCAGTACCAAAATTTTCATGTCAGTACCAAGACAG	1867
Qy	2449	TAGTCAATTCGTGTGGACATAGTTTGGAAACAAAGTCCACAGATATCTGAGGATCTTG	2508
Db	1868	TAGTCAATTTGTGGGTATAGTCTGGAACCAAAATCTTACAAAGTGTCTGAGGACCTTG	1927
Qy	2509	TAAGCATACATCGCCACTCTCTAGAACCTTCTGCTGGTCTTCATGTACGTTTAAAGAGGC	2568
Db	1928	TAAGCATACACCTGCCACTCTCTAGAACCTTCTGCTGGTCTTCATGTACGTTTAAAGCAGAC	1987
Qy	2569	TGGGTGCTGTTTCAAGACTGCATGAATTTGTGTCTTTTTCAGGACTTTTCAAGTAGAGGTAC	2628
Db	1988	TAGGTGCTATTTTCAAGACTGCATGAATTTGTGTCTTTTTCAGGACTTTTCAAGTAGAGGTCC	2047
Qy	2629	TAGTGAATATCTTTACCTGCTGCTGGTGTGGTTCGCCAGGTTGTTGCTGAGATCTGSC	2688
Db	2048	TGGTGGATGACCGCTGCCCTGCCCTGCTGGTGTGCTGAGTGTGCTGAGATGTGGC	2107
Qy	2689	GAAGAAATGGACTGCTCTTATTAGCCAGGTGTTTATTACCAAGATGTTAAAGTGCAGAG	2748
Db	2108	GAAGAAACGGCTCTCACTCATCAGCCAGGTTTCTATTATCAAGATGTTAAATGCAGGG	2167
Qy	2749	AAGAAATGATGATAAAGATATCATCATGCTTCAGATGCTGATGCTTTTAAATGGATCCCA	2808
Db	2168	AGGAAATGTACATAAAGATATCATCATGCTTCAGATGCTGATGCTTTTAAATGGATCCCA	2227
Qy	2809	ATAAGTTCTTGTACTGGTACTTCAGAGTATGAACCTTCCGAGGCTTTTAAACAACCA	2868
Db	2228	ACAAGTTCTTGTACTGGTACTTCAGAGATATGAACCTTACTGATGCTTTTAAACAAGACCA	2287
Qy	2869	TATCTACAAAACACAGGATTTGATTTAAACAATATAATACACTTAATAGAGAAATGCTTC	2928
Db	2288	TATCTACAAAACACAGGATTTGATTTAAACAATATAATACACTTAATAGAGAAATGCTTC	2347
Qy	2929	AGGTCTCTATCTATTTGTGGGTGAGCGTTATGTACTCGGAGTGGGAAATGTGACCAAAAG	2988
Db	2348	AGGTCTCTATCTATTTGTGGGAGAACGTTATGTACTCTGGAGTGGGAAATGTTACCAGAG	2407
Qy	2989	AAGAGTCACAATGAGAGAAATCATTTCACTTTTGCATTTGCAATGCAACCCATGCCACAGTG	3048
Db	2408	AGGAGTTTATAATGAGAGATTTACTCACTTTTGCATTTGCAATGCAACCCATGCCACAGTG	2467
Qy	3049	CCATTGCCAAAAATTTACCTCAGATGAAATTAATGAAACTGGCTTAGAATGTCATAA	3108
Db	2468	CCATTGCCAAAAATTTACCTCAGATGAAATTAATGAAACTGGCTTAGAATGTCATAA	2527
Qy	3109	ACAAAGTGGCCACATTTAAGAAACAGGTTATCAGGCGCATGAGCTTTATGAACTAAAAG	3168
Db	2528	ACAAAGTGGCCACATTTAAGAAACAGGTTATCAGGCGCATGAGCTTTATGAACTAAAAG	2587
Qy	3169	ATGAATCAGCTGAAAGACTTCAATATGTACTTTTATCTACTTCCAAAAACCCAGCATAGCA	3228
Db	2588	ATGAATCAGCTGAAAGACTTCAATATGTACTTTTACCAATATTCTTAAAAACACAGCATAGCA	2647
Qy	3229	AGGCTGAACATATGCAAGAAAGGAGAAACAAAGAAACAAAGATGAAGATTTGCCGC	3288
Db	2648	AGGCTGAACATATGCAAGAAAGGAGAAACAAAGAAACAAAGATGAAGATTTGCCGC	2707
Qy	3289	CACCACCCTCTCTGAAATTTGCGCTCTGCTTTTCAGCAAGTGAATTAACCTTCTCAACTGTG	3348
Db	2708	CGCACCTCTCTCAGAGTTTCTGCCCTGCTTTTCAGCAAGTGAATTAACCTTCTCAACTGTG	2767
Qy	3349	ATATCATGATGTACATTTCTCAGGACCGTATTTTGAGGGGCAATAGACACAGATTTCTAACT	3408
Db	2768	ATGTTATGATATACATCTCTCAGGACCATCTTTTGAGGGGCAATAGACACAGATTTCTAACT	2827
Qy	3409	TGTGACCGAAGGATGCTCCAAATGGCTTTTTCATATTTCTGGCATTTGGTGTTCAGTAAG	3468
Db	2828	TGTGACAGAAGGATGCTTCAGATGCGTTCCTATATATTTGGCATTTGGCTTGTCTGGAAG	2887
Qy	3469	AGAGCAACAGCTTCAAAAAGCTCTGGAAGAGAGTAACATTTTACATTTTATCATAGG	3528
Db	2888	AGAGCAACAGCTTCAAAAAGCTCTGGAAGAGAGTAACATTTTACATTTTATCATAGG	2947
Qy	3529	CTTCAAGATTGGGAAGTTTCAAGCAT-----GAATATACAAATGCTTTTGGAAAAAC	3579
Db	2948	CTTCAAGATTGGGAAGTTTCAAGCATGATGCTCAGATATACAAATGCTCTTGGAAAGAC	3007
Qy	3580	TCAAAGAAATTTCCCAAGTTTGAAGCCAGAGGACATGATAACGTTGGATTTTCAAGATGT	3639
Db	3008	TCAAAGAAATTTCCCAAGTTTGAAGCCAGAGGACATGATAACATGTTGATTTCCAGATGT	3067
Qy	3640	TTGACACAGTCAAGCGATTAAAGAAAAATCTTTTAAATTTCTAGCAACCCATCAGAT	3699
Db	3068	TTGACACAGTCAAGCGATTAAAGAAAAATCTTTTAAATTTCTAGCAACCCATCAGGAC	3127
Qy	3700	CGGAATCTATTAAAGATGATGAGATTACTCATGATAAAGAAAAAGCAGAAAGAA	3759
Db	3128	TGAGTGCATTAAGATGAGGAGATTACTCATGATAAAGAAAAAGCAGAAAGAA	3187
Qy	3760	AAGCTGAAGTGTAGGCTACATGCCAGAAAGATCATGGCTCAGATGCTGCTTTACAGA	3819
Db	3188	AAGCTGAAGTGTAGGCTACATGCCAGAAAGATCATGGCTCAGATGCTGCTTTACAGA	3247
Qy	3820	AAAATCTTCAATGAAACTCAATAAATCATGATGACATAATCATGAGAAATGCTTGGGAAG	3879
Db	3248	AAAATCTTCAATGAAACTCAATAAATCATGATGACATAATCATGAGAAATGCTTGGGAAG	3307
Qy	3880	AAGATTCATTTATGGAGAGAGAGACCCACAGAGTCCAGTCTACTTCTAGAAATTTGCTT	3939
Db	3308	AAGATTCATTTATGGAGAGAGAGACCCACAGTCTAGGAGTCTTCTAGAAATTTGCTT	3367
Qy	3940	TGGTCTCTAAACGGGGTCTTCTTACTTAAAAAGGAGGTGCTGACGTCGATCTCTTCC	3999
Db	3368	TGGGCTCTAAACGGGGTCTTCTTACTTAAAAAGGAGGTGCTGACGTCGATCTCTTCC	3427
Qy	4000	AAGAAGAACAGGAGTGAATAATAGAAATTAATGCCATGATTTATTCGGCTGTGTCCAGA	4059
Db	3428	AAGAAGAACAGGAGTGAATAATAGAAATTAATGCCATGATTTATTCGGCTGTGTCCAGA	3487

Db 5648 TTCTGCTAATATTTCCAGAAACTCTTTCCCTTCATACTGCTCTAGTTCAITTCATAT 5707
QY 6221 CATCCACTTGGTAGAGTAAGTACGACGACGAGTGTAGACATTTTATGTGTTGGTTAAC 6280
Db 5708 CACCCACCTGGTGAATAGAGTCACATTAAGCATTTGTGGACATTTCTCCCATCTGGCTAAC 5767
QY 6281 TCTTCTGCAATTTTGTATTGGTGT 6307
Db 5768 ATCTCTGCACCTTGTATTGGTGT 5794

RESULT 4

AAH14878
ID AAH14878 standard; cDNA; 2550 BP.

XX AAH14878;

XX 26-JUN-2001 (first entry)

XX Human cDNA sequence SEQ ID NO:12731.

DE Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

KW Homo sapiens.

OS EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isoqai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -

XX Claim 8; SEQ ID 12731; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX Sequence 2550 BP; 830 A; 454 C; 558 G; 708 T; 0 other;
SQ
Query Match 40.4%; Score 2546.8; DB 22; Length 2550;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2548; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1189 ATCATGACCTGGAGAGCAGGTACTATAAAGAGAAATTCACGCTGTCGGTTGAATGAAG 1248
Db 1 ATCATGACCTGGAGAGCAGGTACTATAAAGAGAAATTCACGCTGTCGGTTGAATGAAG 60
QY 1249 AGGTAATGTCCAAAGCCAGGAAATATTTCCCTTCAGTGATAAAATATGCTGAGAAATGA 1308
Db 61 AGGTAATGTCCAAAGCCAGGAAATATTTCCCTTCAGTGATAAAATATGCTGAGAAATGA 120
QY 1309 CTATATGGGAAGAGGAAAAAGAACTGCTCCTCGAACTCCAGATAAGGGAGAAAAATGAAA 1368
Db 121 CTATATGGGAAGAGGAAAAAGAACTGCTCCTCGAACTCCAGATAAGGGAGAAAAATGAAA 180
QY 1369 GATACTATTGTGCTCTTTTCAATGATGAACACCATTCATATGACCACGTCATATACAGCC 1428
Db 181 GATACTATTGTGCTCTTTTCAATGATGAACACCATTCATATGACCACGTCATATACAGCC 240
QY 1429 TACAAAGAGCTCTTGACTGTGAGCTCGCAGAGGCCAGTTGCATACCACTGCCATTGACA 1488
Db 241 TACAAAGAGCTCTTGACTGTGAGCTCGCAGAGGCCAGTTGCATACCACTGCCATTGACA 300
QY 1489 AAGAGGCTGCTGGGCTGTTAAAGCGGAGGCTTATGCTGCTGCAGGAGCAAGGAAG 1548
Db 301 AAGAGGCTGCTGGGCTGTTAAAGCGGAGGCTTATGCTGCTGCAGGAGCAAGGAAG 360
QY 1549 ATATAAGAGTCATTCAGAAAATGCTCTCAACATCCACTTCATGTAGAGTATTACACT 1608
Db 361 ATATAAGAGTCATTCAGAAAATGCTCTCAACATCCACTTCATGTAGAGTATTACACT 420
QY 1609 CAGAGATTATGGCTCATCAGAAATTTGCTTCCGCTCTTGGTTCCTGGATGAACAAATTA 1668
Db 421 CAGAGATTATGGCTCATCAGAAATTTGCTTCCGCTCTTGGTTCCTGGATGAACAAATTA 480
QY 1669 TGAGCTATTCAAGTGACTTTAGGCAGATCTTTTGGCAAGCATGCTCTAGAGAAAGACCTG 1728
Db 481 TGAGCTATTCAAGTGACTTTAGGCAGATCTTTTGGCAAGCATGCTCTAGAGAAAGACCTG 540
QY 1729 ACTCGGAGATCCCTGCTCTCATAGCAGGTTAATGCTTTGGGATGCAAAAGCTTTATPAAG 1788
Db 541 ACTCGGAGATCCCTGCTCTCATAGCAGGTTAATGCTTTGGGATGCAAAAGCTTTATPAAG 600
QY 1789 GTGCCGTAAGATCCTTCATGAATGTATCTTCAGCAGGTTTTTTTATGGAGATGGAATACA 1848
Db 601 GTGCCGTAAGATCCTTCATGAATGTATCTTCAGCAGGTTTTTTTATGGAGATGGAATACA 660
QY 1849 AAAAATCTTTGCTATGCAATTTGGAAGTATTATAAACAACCTGAGAGAAAGATATATCA 1908
Db 661 AAAAATCTTTGCTATGCAATTTGGAAGTATTATAAACAACCTGAGAGAAAGATATATCA 720
QY 1909 GTGATGATCATGACAGAGATGATCTCTAATGCTGCACTTTTCAGTTCAGATGTTTACTGTTC 1968
Db 721 GTGATGATCATGACAGAGATGATCTCTAATGCTGCACTTTTCAGTTCAGATGTTTACTGTTC 780
QY 1969 CTACTCTGGCTCGACATCTTATTGAAGAGCAGAAATGTTATCTCTGTCATTAAGAACTC 2028
Db 781 CTACTCTGGCTCGACATCTTATTGAAGAGCAGAAATGTTATCTCTGTCATTAAGAACTC 840
QY 2029 TGCTAGAGATTTTACCTGAGTACTTGGACAGGAAACAATAAATTCACCTCCAGGGTTATA 2088
Db 841 TGCTAGAGATTTTACCTGAGTACTTGGACAGGAAACAATAAATTCACCTCCAGGGTTATA 900
QY 2089 GCCAGGACAAATTTGGGAAGAGTATATGCAGTAATATGTGACCTAAAGTATATCCCTGATCA 2148
Db 901 GCCAGGACAAATTTGGGAAGAGTATATGCAGTAATATGTGACCTAAAGTATATCCCTGATCA 960
QY 2149 GCAAAACCCACCAATATGGACAGAAAGATTAAGAAATCAGTTCCTTGAAGGTTTTTCGATCTT 2208
Db 2149 GCAAAACCCACCAATATGGACAGAAAGATTAAGAAATCAGTTCCTTGAAGGTTTTTCGATCTT

Db	961	GCAACCCACATATGACAGAAAGATTAGAATGCAGTTCCTTGAAGGTTTTCATCTT	1020
Qy	2209	TTTTGAAGATCTTACCTGTATGCAGGAATGGAAGAAATCCGAAGACAGGTGGGCAAC	2268
Db	1021	TTTTGAAGATCTTACCTGTATGCAGGAATGGAAGAAATCCGAAGACAGGTGGGCAAC	1080
Qy	2269	ACATTGAAGTGGATCCTGATTGGAGGCTGCCATTGCTATACAGATGCAATTTGAAGATA	2328
Db	1081	ACATTGAAGTGGATCCTGATTGGAGGCTGCCATTGCTATACAGATGCAATTTGAAGATA	1140
Qy	2329	TTTTTACTCATCTTCCAAAGTGGTGTCTTGTGATGAAGAACTCTTACTTGTGGCTTATA	2388
Db	1141	TTTTTACTCATCTTCCAAAGTGGTGTCTTGTGATGAAGAACTCTTACTTGTGGCTTATA	1200
Qy	2389	AAGAATGTCACAAAGCTGTGATGAGTGCAGTACCAGTTTCATATCTAGTAGCAAGACAG	2448
Db	1201	AAGAATGTCACAAAGCTGTGATGAGTGCAGTACCAGTTTCATATCTAGTAGCAAGACAG	1260
Qy	2449	TAGTACAATCGTGTGGACATAGTTTGGAAACAAAGTCCCTACAGAGTATCTGAGGATCTTG	2508
Db	1261	TAGTACAATCGTGTGGACATAGTTTGGAAACAAAGTCCCTACAGAGTATCTGAGGATCTTG	1320
Qy	2509	TAAGCATACATCTGCCACTCTCTAGGACCCCTTGTGCTCTTCATGTAGTCTTAAGCAGGC	2568
Db	1321	TAAGCATACATCTGCCACTCTCTAGGACCCCTTGTGCTCTTCATGTAGTCTTAAGCAGGC	1380
Qy	2569	TGGTGTCTGTTTCAAGACTGCATCAATTTGTGCTCTTTGAGGACTTTCAGTAGAGGTAC	2628
Db	1381	TGGTGTCTGTTTCAAGACTGCATCAATTTGTGCTCTTTGAGGACTTTCAGTAGAGGTAC	1440
Qy	2629	TAGTGGAAATCCCTTACGTTGTCTGTGGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGT	2688
Db	1441	TAGTGGAAATCCCTTACGTTGTCTGTGGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGT	1500
Qy	2689	GAGAAATGGACTCTCTTATTAGCCAGGTGTTTTATTACCAAGATGTTAAGTGCACAG	2748
Db	1501	GAGAAATGGACTCTCTTATTAGCCAGGTGTTTTATTACCAAGATGTTAAGTGCACAG	1560
Qy	2749	AAGAAATGTATGATAAAGATATCATCTGCTTCAGATGGTGGCATCTTTAATGGATCCCA	2808
Db	1561	AAGAAATGTATGATAAAGATATCATCTGCTTCAGATGGTGGCATCTTTAATGGATCCCA	1620
Qy	2809	ATAAGTCTTGTACTGTTACTGAGGATATGAACCTTGCAGGCTTTTAAACAGACCA	2868
Db	1621	ATAAGTCTTGTACTGTTACTGAGGATATGAACCTTGCAGGCTTTTAAACAGACCA	1680
Qy	2869	TATCTACAAAGACAGGATTTGATTAACAATATATACACTAATAGAAAGAAATGCTTC	2928
Db	1681	TATCTACAAAGACAGGATTTGATTAACAATATATACACTAATAGAAAGAAATGCTTC	1740
Qy	2929	AGTCTCATCTATATTGTGGTGAGCGTGTATGTACCTGGAGTGGGAAATGTGACCAAG	2988
Db	1741	AGTCTCATCTATATTGTGGTGAGCGTGTATGTACCTGGAGTGGGAAATGTGACCAAG	1800
Qy	2989	AAGAGGTACAAATGAGAGAAATCATCTTCTGTTGCAITTTGAACCCCATGCCACACAGTG	3048
Db	1801	AAGAGGTACAAATGAGAGAAATCATCTTCTGTTGCAITTTGAACCCCATGCCACACAGTG	1860
Qy	3049	CCATTGCCAAAATTTACCTGAGAAATGAATAATATGAACCTGGCTTAGAGAAATGTCATAA	3108
Db	1861	CCATTGCCAAAATTTACCTGAGAAATGAATAATATGAACCTGGCTTAGAGAAATGTCATAA	1920
Qy	3109	ACAAAGTGGCCACATTTAAGAAACAGGTGTATCAGGCCATGGAGTTTATGAACATAAAG	3168
Db	1921	ACAAAGTGGCCACATTTAAGAAACAGGTGTATCAGGCCATGGAGTTTATGAACATAAAG	1980
Qy	3169	ATGAATCAGTCAAGACTTCAATATGTACTTTTATCTTACTCTCCAAAACCCAGCATAGCA	3228
Db	1981	ATGAATCAGTCAAGACTTCAATATGTACTTTTATCTTACTCTCCAAAACCCAGCATAGCA	2040
Qy	3229	AGGCTGAACATATGCAAGAAAAAGGAGAAAAACAAAGATAAGATGAAGCATTTGCCGC	3288
Db	2041	AGGCTGAACATATGCAAGAAAAAGGAGAAAAACAAAGATAAGATGAAGCATTTGCCGC	2100

Qy	3289	CACCACCACCTCTCGAATTTCTGCCCTGCTTTTCAGCAAAAGTGAATTAACCTTCTCAACTGTG	3348
Db	2101	CACCACCACCTCTCGAATTTCTGCCCTGCTTTTCAGCAAAAGTGAATTAACCTTCTCAACTGTG	2160
Qy	3349	ATATCATGATGTACATTTCTCAGGACCGTATTTTGAGCGGCAATAGACAGATTTCTAACT	3408
Db	2161	ATATCATGATGTACATTTCTCAGGACCGTATTTTGAGCGGCAATAGACAGATTTCTAACT	2220
Qy	3409	TGTGACCGCAAGGGATGCTCCAAATGGCTTTTCATATCTGGCATTTGGTCTTACTAGAAG	3468
Db	2221	TGTGACCGCAAGGGATGCTCCAAATGGCTTTTCATATCTGGCATTTGGTCTTACTAGAAG	2280
Qy	3469	AGAACAACAGCTTCCAAAAGCTCCTGAAAGAAAGTAACAATTTTGACTTTTATCATAAGG	3528
Db	2281	AGAACAACAGCTTCCAAAAGCTCCTGAAAGAAAGTAACAATTTTGACTTTTATCATAAGG	2340
Qy	3529	CTTCAAGATTGGGAAGTTTCAGCCATGAATATACAAATGCTTTTGGAAAACTCAAAGGAA	3588
Db	2341	CTTCAAGATTGGGAAGTTTCAGCCATGAATATACAAATGCTTTTGGAAAACTCAAAGGAA	2400
Qy	3589	TTCCCCAGTTAGAAGCCAGAGGACATGATAACGTGGATACTTTCAGATGTTTGACACAG	3648
Db	2401	TTCCCCAGTTAGAAGCCAGAGGACATGATAACGTGGATACTTTCAGATGTTTGACACAG	2460
Qy	3649	TGAACGATTAAGAGAAAAATCTTGTTTAAATTTGTAGCAACCATCAGGATCGGAATCTA	3708
Db	2461	TGAACGATTAAGAGAAAAATCTTGTTTAAATTTGTAGCAACCATCAGGATCGGAATCTA	2520
Qy	3709	TTAAGAATGATGAGATTTACTCATGATAAG 3738	
Db	2521	TTAAGAATGATGAGATTTACTCATGATAAG 2550	

RESULT 5

ABQ75898
ID ABQ75898 standard; cDNA; 1635 bp.

XX ABQ75898;

DT 17-OCT-2002 (first entry)

XX Human ubiquitin relative protein 46.64 cDNA.

DE Human; ubiquitin relative protein 46.64; tumour; inflammation;
KW immunological disease; haemopathy; human immunodeficiency virus; HIV;
KW gene, ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 55..1329

FT /tag- a

FT /product= "ubiquitin relative protein 46.64"

XX CN1339485-A.

PD 13-MAR-2002.

PF 23-AUG-2000; 2000CN-0119708.

PR 23-AUG-2000; 2000CN-0119708.

XX (BODE-) BODE GENE DEV CO LTD SHANGHAI.

XX Mao Y, Xie Y;

XX WPI: 2002-464069/50.

DR P-PSDB; ABQ75898.

XX New polypeptide-human ubiquitin relative protein 46.64 for
PT treating malignant tumours, inflammations, immunological diseases,
PT haemopathy and human immunodeficiency virus infection -

XX
PS
XX

Claim 6; Page 24-25 (disclosure); 33pp; Chinese.

The present invention discloses a new kind of polypeptide, human ubiquitin relative protein 46.64, polynucleotides for encoding this polypeptide, and a DNA recombination process to produce the polypeptide. The present invention also discloses the method of applying the polypeptide in treating various diseases, such as malignant tumours, inflammations, immunological diseases, haemopathy and human immunodeficiency virus (HIV) infection. The current sequence represents the human ubiquitin relative protein 46.64 cDNA.

XX
SQ .Sequence 1635 BP; 529 A; 290 C; 344 G; 472 T; 0 other;

Query Match 20.8%; Score 1311; DB 24; Length 1635;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 1317; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY	2127	GACCTAAGTATATCTGATCAGCAACCCACACATATGGACAGAAATTAAGAAATGCAG	2186
DB	1	GACCTAAGTATATCTGATCAGCAACCCACACATATGGACAGAAATTAAGAAATGCAG	60
QY	2187	TTCTTTGAAGTTTTCGATCTTTTGAAGATCTTACCTGTATGCAGGAATGGAAGAA	2246
DB	61	TTCTTTGAAGTTTTCGATCTTTTGAAGATCTTACCTGTATGCAGGAATGGAAGAA	120
QY	2247	ATCCGAAGACAGTTGGCAACACATTTGAAGTGCATCTGATTTGGAGGCTGCCATTGCT	2306
DB	121	ATCCGAAGACAGTTGGCAACACATTTGAAGTGCATCTGATTTGGAGGCTGCCATTGCT	180
QY	2307	ATACAGATGCAATTTGAAGATATTTTACTCATGTTTCCAAAGTGGTGTGTGTGATGAA	2366
DB	181	ATACAGATGCAATTTGAAGATATTTTACTCATGTTTCCAAAGTGGTGTGTGTGATGAA	240
QY	2367	GAACTCTTACTTGTGCTTATAAGAAATGTCACAAAGCTGTGATGAGTGTCAGTACCAGT	2426
DB	241	GAACTCTTACTTGTGCTTATAAGAAATGTCACAAAGCTGTGATGAGTGTCAGTACCAGT	300
QY	2427	TTCATATCTAGTAGCAAGACAGTAGTACAAATCGTGTGGACATAGTTTGGAAACAAAGTCC	2486
DB	301	TTCATATCTAGTAGCAAGACAGTAGTACAAATCGTGTGGACATAGTTTGGAAACAAAGTCC	360
QY	2487	TACAGAGTATCTGAGGATCTTGAAGCATACATCTGCCACTCTCTAGGACCCCTTGCTGT	2546
DB	361	TACAGAGTATCTGAGGATCTTGAAGCATACATCTGCCACTCTCTAGGACCCCTTGCTGT	420
QY	2547	CTTCATGTAGTTTAAAGCAGCTGGGTGCTGTTTCAAGACTGCATGAATTTGTCTTTT	2606
DB	421	CTTCATGTAGTTTAAAGCAGCTGGGTGCTGTTTCAAGACTGCATGAATTTGTCTTTT	480
QY	2607	GAGGACTTTCAAGTAGAGGTACTAGTGGAAATATCCTTTAGCTGTCTGTGTTGGTTGCC	2666
DB	481	GAGGACTTTCAAGTAGAGGTACTAGTGGAAATATCCTTTAGCTGTCTGTGTTGGTTGCC	540
QY	2667	CAGGTTTGTCTGAGATGTGGCGAAGAAATGGACTGTCTTATAGCCAGGTGTTTAT	2726
DB	541	CAGGTTTGTCTGAGATGTGGCGAAGAAATGGACTGTCTTATAGCCAGGTGTTTAT	600
QY	2727	TACCAAGATGTTAAGTGCAGAGAAATGTATGATATAAATATCATCTGCTTCAGATT	2786
DB	601	TACCAAGATGTTAAGTGCAGAGAAATGTATGATATAAATATCATCTGCTTCAGATT	660
QY	2787	GGTGCATCTTTAATGGATCCCAATAGTTCTTGTACTGTACTTTCAGAGGTATGAACCTT	2846
DB	661	GGTGCATCTTTAATGGATCCCAATAGTTCTTGTACTGTACTTTCAGAGGTATGAACCTT	720
QY	2847	GCCGAGGCTTTTAAACAGACCATATCTTACAAAAGCAGGATTTGATTAAACATAAAT	2906
DB	721	GCCGAGGCTTTTAAACAGACCATATCTTACAAAAGCAGGATTTGATTAAACATAAAT	780
QY	2907	ACACTAATAGAAGAAATGCTTACAGTCCCTCATCTATTTGGGTGAGCGTTATGTACCT	2966
DB	781	ACACTAATAGAAGAAATGCTTACAGTCCCTCATCTATTTGGGTGAGCGTTATGTACCT	840

QY	2967	GGAGTGGAAATGTGACCAAGAGAGTCACAATGAGAGAAATCATTTCACTTCTGCTTTC	3026
DB	841	GGAGTGGAAATGTGACCAAGAGAGTCACAATGAGAGAAATCATTTCACTTCTGCTTTC	900
QY	3027	ATTGAACCATGCCCACACAGTCCCATTTGCCAAAAATTTACCTTGAGAAATGAAATGAA	3086
DB	901	ATTGAACCATGCCCACACAGTCCCATTTGCCAAAAATTTACCTTGAGAAATGAAATGAA	960
QY	3087	ACTGCTTTAGAGAATGTCTATAAACAAGTGGCCACATTTTAAGAAAACAGGTGTATCAGG	3146
DB	961	ACTGCTTTAGAGAATGTCTATAAACAAGTGGCCACATTTTAAGAAAACAGGTGTATCAGG	1020
QY	3147	CATGAGTTTATGAACATAAGATGAATCACTGAAAGACTTCAATATGTACTTTTATCAT	3206
DB	1021	CATGAGTTTATGAACATAAGATGAATCACTGAAAGACTTCAATATGTACTTTTATCAT	1080
QY	3207	TACTCCAAAACCCAGCATAGCAAGCTGAACATATGCAGAAGAAAAGGAGAAAACAGAA	3266
DB	1081	TACTCCAAAACCCAGCATAGCAAGCTGAACATATGCAGAAGAAAAGGAGAAAACAGAA	1140
QY	3267	AACAAAGATGAAGCATTTGCCGCCACACACCTCTCTGAATTTCTGCCCTGCTTTTCAGCAA	3326
DB	1141	AACAAAGATGAAGCATTTGCCGCCACACACCTCTCTGAATTTCTGCCCTGCTTTTCAGCAA	1200
QY	3327	GTGATTAACCTTCTCAACTGTGATATCATGATGTACATTTCTCAGGACCCGTATTTGAGCGG	3386
DB	1201	GTGATTAACCTTCTCAACTGTGATATCATGATGTACATTTCTCAGGACCCGTATTTGAGCGG	1260
QY	3387	GCAATAGACACAGATTCTTAACCTTTGGACCCGAAGGATGCTCCAAATGCTTTTCATATT	3446
DB	1261	GCAATAGACACAGATTCTTAACCTTTGGACCCGAAGGATGCTCCAAATGCTTTTCATATT	1320
QY	3447	CTGGCAT 3453	
DB	1321	CAGGCAT 1327	

RESULT 6
AAK51709
ID AAK51709 standard; cDNA; 6840 BP.
XX AAK51709;
AC AAK51709;
DT 06-NOV-2001 (first entry)
XX Human polynucleotide SEQ ID NO 254.
DE Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.
XX Homo sapiens.
XX WO200157190-A2.
XX 09-AUG-2001.
XX 05-FEB-2001; 2001WO-US04098.
PF 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX (HYSE-) HYSEQ INC.
PA Tang Yt, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI

Db 1897 TAACATTTTGCAGTCAGATGTCATGTTGTGTCATCATGGGAACAATTTCTGCAATGGCGTGT 1956
Qy 3392 AGACACAGATTTCTAACTTGTGGACCGAAGGATGCTCCAAATGGCTTTTCATATTCGGC 3451
Db 1957 GGAACATAATGGATATGCCCTGGTCAGAGTCCTGCAAGGGGTGTACATTTAAATGG 2016
Qy 3452 ATTGGGTTTACTAGAAGAGAACACAGCTTCAAAAAGCTCCTGAAG---AAGAAGTAAC 3508
Db 2017 CATGGCATTACAAGAGAAAACAACATTTAGAGATGTCAAGGAGCATGTAGTAAC 2076
Qy 3509 ATTGGACTTTTATCAATAGGCTTCAAGATTTGGGAAGTTTCAGCCCATGAAT-----AT 3559
Db 2077 ATTTAGCTTTCACATCAAGAGATATCAAAACCTGTGTGAAGCGCCAAAATAATTTCTCCTAGCAT 2136
Qy 3560 ACAATAGCTTTTGGAAAACTCAAGGAATTTCCCGAGTTAGAGGCCAGAGACATGAT 3619
Db 2137 ACTAGCTATGCTGGAACATCAAAAATGCTCCCTACTAGAAGTCACAAAGACATGAT 2196
Qy 3620 AAGCTGGATACTTTCAGATGTTTGACACAGTGAAGCGATTAAGAGAAAAATCTTCTTTAAT 3679
Db 2197 TCGGTGGATATGAAGACTTTTAAATGCTGTATAAAGATGAGGGAGA-----GTTCAAC 2250
Qy 3680 TGTAGCAACACATCAGGATCGGAATCTATTAAAGATGATGAGATTACTCATGATAAGA 3739
Db 2251 TACCAGTCCGTCGACAGACAGAAGAACCATAAATGGAAGAGATTTCAAGGGACAAGA 2310
Qy 3740 AAAGCAGAACGAAAAAGAAAGCTGAAGCTGCTAGGCTACATCGCCAGAGATCATGCG 3799
Db 2311 CAAAGCTGAGAGGAAGAAAGAACAGAGATTTGCCAGATCGCAGAGAAAAAGATCATGCG 2370
Qy 3800 TCAGATGCTCGCCTTACAGAAAACTTCATTGAACTCATAACTCATGATGACATAC 3859
Db 2371 TCAGATGCTGAAATCAGCGGATTTTATTGATGAAACAAAGAACTCTTTACGACAG 2430
Qy 3860 ATCAGAAATGCCCTGGGAAAGAAATTCATTTATGGAGGAAGAGACACCCAGCAGTCAG 3919
Db 2431 ATTAGAATGGATGCGCTCAACCTCTGCTGTCTTGATCAI-----AGCCCTGTGGCTC 2484
Qy 3920 TGACTACTTAGAATTTGCTTTGGTCTTAACGGGGTCCATCTGTCTTACTGAAAGGAGGT 3979
Db 2485 AGATATGACACTTTACAGACATGCGGCCCGCACAACTCAGGTCTCCGTAACAAAGACAAT 2544
Qy 3980 GCTGACGTGCATCCTTTGGCAAGAACAGAGAGGTGAATAAGAAATATGCCATGCT 4039
Db 2545 CGTTACATGATATTGTGTCAAGAGGAGCAAGAGTTAAAGTGGAAAGCGGGCAATGGT 2604
Qy 4040 ATTATCGGCTGTGTCCAGAAATCTACTGCTTAAACCCAGCACAGGGGAAACCCATAGA 4099
Db 2605 CTGGCAGCATTTGTTTCAGAGATCAACTGTATTATCAAAAACAGAGTAAATTTATCA 2664
Qy 4100 ACTCTCAGAGAAAGCCCTAGACCCACTTTTCATGGATCCAGACTTGGCATATGGAACCTA 4159
Db 2665 AGATCCAGAAAA---TATGATCCATTATTCATGACCCCTGATCTGTGTGTGGAACACA 2721
Qy 4160 TACAGAAAGCTGTGGTCATGTAATGACGCAAGTGTGCTGGCAGAGTATTTTGAAGCTGT 4219
Db 2722 CACTAGTAGCTGTGGGCATATTATGATGCCCCATTTGTTGCAAGAGTATTTTGAATCCGT 2781
Qy 4220 ACAGCTGAGCTCTCAGCAGGCAATTCATGTTGACCTTTT-----TGACTTGA 4267
Db 2782 TCAGCTAAGAACAGGAGGCAACAGAGATTACGCTTACATACGAGCTATGATGTAGA 2841
Qy 4268 AAGTGAGATATCTTTGGCCCTTTTGGCAATCTCTGCAATCTGTGCAATCTGTATCCCATTTAT 4327
Db 2842 AAACGAGAAATTTCTTTGGCCCTTTTGTGAATGCTTGAGTAATACCTGTTATTC----- 2895
Qy 4328 TCTTTGCAACCTCAAAAGATAAACAGTGCAGATGCTCTTGTGCTCAACTTTTGCAC 4387
Db 2896 TCTGCTGCTTCTCCCAAGAAATATTTTAAACAACAGGTTAAATTTTTCAGACCAACCAA 2955
Qy 4388 CTGGCAGGCTGGATACAGACTGTTCTGGCCAGAAATATCAGGTTATAATATAGACATGC 4447
Db 2956 TCTGACTCAGTGGATTAGAACATAATCTCAGCAAAATAAAAGCATTTACAGTTTCTTAGGAA 3015

Qy 4448 TAAAGGAGAAACCCAAATTCCTATTCTTTTAATCAAGGAATGGGAGATTTCTACTTTGGA 4507
Db 3016 AGAAGAAAGTACTCTCTAAATATGCGCTCTACAAAGAAATTCAGAAAATGTGGATGAATACA 3075
Qy 4508 GTTCCATTTCCATCCTGAGTTTTTGGGCTTGAGTCTTCGATTTAAATATTAATTAAGATGACATCA 4567
Db 3076 GCTCCTGAAGGTTTCAGGCCCTGATTTTCGTCTTAAGATCCCTTATTTCTGAGAGCATAAA 3135
Qy 4568 GGAATGTTTATTTCTTTTGGCCACAAACAATTTATAGAAATGGATTTGAAAGTGGCCACTGA 4627
Db 3136 AGAATGCTAACGACATTTTGGAACTGCTACCTACAAGGTGGGACTAAAGGTTTCATCCCAA 3195
Qy 4628 TGAAGGATCCTTCAGTCCCATGCTGACCTGGAGCACCTCGGCTTTTCACTATCCAGGC 4687
Db 3196 TGAAGAGATCCTCGTGTTCCTCAATATGTTTGGGGTAGCTGCGGTACACCATCCAAAG 3255
Qy 4688 AATTGAAAATCTATTGGGAGATGAAGGAAACCTCTGTTTGGAGCACTTTCAAAATAGGCA 4747
Db 3256 CATAGAAAGAAATTTGAGTGATGAGATAAACCATTTGTTGGTCTTTACCTTGCAGACT 3315
Qy 4748 GCATAATGCTCTGAAAGCATTAATGCAGTTTGCAGTTGCACAGAGATTACCTTGTCTCA 4807
Db 3316 GGATGACTGTCTTAGTTCATTGACGAGATTTGCGCGACGACACTGGACAGTGGCATCAGT 3375
Qy 4808 GGTCTGATACAGAAACATCTGGTTCCTTCTATCAGTTGTTCTTCTTAACATAAATC 4867
Db 3376 TTCAGTGTGGAAGGACATTTTGTAAACCTTTTGATCAGTCTGCTTAATGACAGCCA 3435
Qy 4868 AGAAGATACACCATGCTCTGCTATAGATGTTTTCATGTTTGGTGGGTGCTGTGT 4927
Db 3436 TGAGGAATTCATGATATAGATGTTTTCATTTATTTATTTGTTGGGTGTTGTTGCT 3495
Qy 4928 AGCATTCCTCATCTTGTATTGGATGACCTCTGATCTGACGCTTCTTTCAGTTAGTTTC 4987
Db 3496 TGCAATTCCTGGTTGCAGTG-----TCAGGATTTTCAGGGATCAGCCT 3540
Qy 4988 TTCTATAAACACCTTTTATCTCTTCCATTTGATACCAATGACGACACACTGCTTCAGATACT 5047
Db 3541 TGGCACTGGAGACCTTTCACATTTTCCATCTGCTTACTATGGCACACATCATACAGATCTT 3600
Qy 5048 ACTTACAGTAGACACAGGCTTACC---CCTTCTCAGTTCAAGAGACAGACAGGAGGC 5104
Db 3601 ACTTACCTCATGTACAGAAAGAAATGGCATGATCAAGAAAATFCCCTTTGTGAAGAAGA 3660
Qy 5105 TCATTCCCATCTTCTTTTTCAGAAATTTCTCAATATACAAGTGGCTCCATTTGGTG 5164
Db 3661 ATCAGCAGTTCTTGTATTAACACTTCACCAGTATACGGGAGTGCCTTTGAAAGA 3720
Qy 5165 TGATATTCTCGCTGGTATTTGTGGGTCTACTGAAGAATGGCATCACCCCTTATCTTCG 5224
Db 3721 AATACCATCCGCTGGCATCTGTGGAGGAGTGTGAGAGCTGGAATCATGCTTTCTCTGAA 3780
Qy 5225 CTGTGCTCATTTGTTTTCCTACTATTTACTTGGGTAACTCGCCTGAGAACTGCATAC 5284
Db 3781 GTGTTCTGTTATTTTTCATTTACTTAAATGGAGTTCTTCCCTCCACCGGAGATTTCAAGT 3840
Qy 5285 CAATTCTCAGAAAGGAGAGTACAGTGCATCTGTAGCTATCTATCTTTACCTACAAATTT 5344
Db 3841 ---TCTTGGAAAGCCCAATTTTGAACATTTATGTAGTATCTTCTCCCTACCAAAACCT 3897
Qy 5345 GTTCTGCTTTCAGGAATATTGGTACTCTTAAGGGCCCTTTGCTCCAGAGGTGGTGTGC 5404
Db 3898 CATTTGCCCTTTTCAAGAAAATAGTGAGATAATGAATTCACATGATTTGAAAGTTGGTGGCG 3957
Qy 5405 AGATCCTGCCCTTAACTGTTTGAAGCAAAAACACCCGTTGGTGCAGTACCCCTAGAAA 5464
Db 3958 TAACAGTGAAGTAAAAGATATCTAGAAGGTGAAAGAGATGCTATAAGATATATCCAAGAGA 4017
Qy 5465 AAGAAATAGTTTATAGAGCTTCTTGATGACTATAGTGCCTTCTGCTCAAGCTTCTCA 5524
Db 4018 ATCTAACAAATTAATAAACCTTCCAGAGGATTTACAGCAGCTCATTAATCAAGCATCCAA 4077

937	GTCTCGCTGTA	GTGCACTGTC	ATGCTGGTGGT	TTATACTGATG	TGTAACACCC	CAATCACAC	936
2452	TACAATCGT	GTGACATAGT	TTTGGAAACA	AAAGCTCCT	CACAGAGTAT	CTGAGGATCTT	2511
997	TAAGCATTT	TGCGACATTC	AGTGGAAACT	ATCAGATAC	TGTTGTTT	CCCAAGAAAA	1056
2512	GCATACATC	TGCCACACTC	TCTAGNACC	TTTGGCTT	GTCTCATGT	TACGTTTAA	2571
1057	GCATTCACC	TCCGAGTTT	CTCGGTTACT	TCCGAGGTT	TACATGTAT	TATTAA	1116
2572	GTGCTGTTT	CAAGACTGC	ATGAATTTG	TGCTTTTG	AGGACTTCA	AGTAGAGT	2631
1117	AAGTGGCAT	ATAAATTC	ASAGCTCCT	ACTCTA	AGTGAAC	TAGCCCA	1176
2632	TGGAATATC	CTTTACGTT	GTCTGGTG	TGGTGTCC	CCAGGTTG	TGTGTGAGAT	2691
1177	TAGAACACC	CTCTTAGAT	GTCTTGTG	TGTCGCCA	AGTACATG	CCGGAATG	1236
2692	GAATGGACT	GTCTCTTAT	TAGCCAGG	TGTTTTAT	TACCAAGAT	TGTTAAGT	2751
1237	GAATGGGTT	CTCTAGTAA	ACCAGATTT	TATTACT	ACCATAAT	GTGAAATG	1296
2752	AAATGTATG	ATAAGATAT	CATCATG	CTTCAGAT	TGGTGCAT	CTTTAATG	2811
1297	AGATGTTG	ACAGGATGT	AGTAATG	CTTCAGAC	AGGTGCTC	CCATGATG	1356
2812	AGTCTTGT	TACTGGTACT	TTCAGAGG	TATGAAC	TGTCGGAGG	CTTTTAA	2871
1357	ATTTCCGT	ATGATCAT	GCTCAGC	CGCTTTG	AACTTTAT	CAGATTTT	1416
2872	CTACAAA	G-----	-----	-----	-----	-----	2910
1417	ATGGAAA	AGAGATT	TAGTCTG	AGATTAC	CCCATA	AGGATGTG	1476
2911	TAATAGA	AGAAATG	CTTCAGG	TCTCAT	CTATAT	TGTGGGTG	2970
1477	TAATAGA	AGAAATG	CTATAC	TCTCAT	TATATG	TCTGTG	1536
2971	TGGAAAT	GTGACCA	AAAGAG	GGTCA	CAATG	AGAGAAAT	3030
1537	TTGACAG	GGTAAATG	CTACAG	ATGAAAT	CAAGC	GAGAGATT	1596
3031	AACCCAT	GCCACAC	AGTGCC	ATTTGCC	AAAAAT	TTTAC	3090
1597	AGCCTAT	GGCTCAT	AGTGAAT	TGGTA	AGTCTTT	TAC	1656
3091	GCATTAGA	GAATGC	ATAAACA	AAAGTG	CGCCAC	ATTTAG	3150
1657	GCATGGAG	AGTGTAA	TGCGA	AGCTT	TGCCCAT	TTTCA	1716
3151	GAGTTTAT	GAACTAAA	AGATGA	ATCACT	TGAAAG	ACTTCA	3210
1717	GCATGAT	GAACGAA	CCAGAT	GTGCC	AAAGAT	CTCACT	1776
3211	CCAAAA	CCCCAG	CATPAG	CAGGCTG	GAACAT	ATG	3270
1777	CAAGGCG	CAACAG	TCTCC	AAAGC	AGAAG	CGCAAA	1836
3271	AAGATGA	AGCATAT	TGCC	CGCACC	ACCTCCT	CGAAT	3330
1837	AAGATAC	AGCACTCCC	ACCCTCC	GGTGTG	CTGCTCC	ATCTG	1896
3331	TTAACCT	TTCACTG	TATCAT	GTATG	TACAT	CTCTC	3390
1897	TTACAT	TTTGC	AGTCAG	ATGTCA	TGTG	CAAT	1956
3391	TAGAC	CACAGAT	TTTCA	CTTGTG	GACCA	GGGATG	3450
1957	TGGA	CAATAT	TGGAT	TGCTG	TGC	AGTCC	2016
3451	CATTGG	GGTTACT	TAG	AAGAG	CAAC	AGCTT	3507
2017	GCATGG	CATCTAC	AGA	AGAAAA	CAACAT	TTT	2076

QY	3508	CATTGTGACTTTTATCATAAAGCGCTTCAAGATTGGGAAGTTTCAGCCATGAAT-----A	3550
DB	2077	CATTTCCTTCACTCAGAAGATATCAAAAGCCTGGTGAAGCGCCAAAAAATTTCTCTAGCA	2136
QY	3559	TACAAATGCCTTTTGGAAAAAATCAAAGGAATTCGCCAGTTAGAAGGCCAGAAGCATGA	3618
DB	2137	TACTAGCTATGCTGGAAACATACAAATGCTCCCTACCTTGAAGTTCACAAAGACATGA	2196
QY	3619	TAACGTGGATACTTCAGATGTTTGACACAGTGAAGCGATTTAAGAGAAAAATCTTGTTAA	3678
DB	2197	TTGCGTGGATATTGAACACTTTTAAATGCTGTAAAAAGATGAGGGAGA-----GTTTAC	2250
QY	3679	TTGTGAAACACCATCAGGATCGGAATCTATTAAAGATGATGAGATTACTCATGATAAG	3738
DB	2251	CTACCAGTCCGTTGGCAGACAGAGGAACCATTAATGGAAGAGAGTTCAAGGGACAAAG	2310
QY	3739	AAAAAGCAGAAAGAAAAAGAAAGCTCAAGCTGCTAGGCTACATCGCCAGAAGCATG	3798
DB	2311	ACAAGCTGAGAGAGAGAAAGACGAGATTTGCCAGCTGCCGAGAAAGATCATGG	2370
QY	3799	CTCAGATGCTGCTCCCTTACAGAAAAAATCTTATGAAACTCATAAACTCATGTATGACAATA	3858
DB	2371	CTCAGATGCTGAAATCCAGCGGCAATTTATTTGATGAACAAGAAGTCTTTCAGCAGA	2430
QY	3859	CATCAGAAATGCCCTGGGAAAGAGATTCCATTTATGGAGGAAGAGACACCCAGCAGTCA	3918
DB	2431	CATTAGAACGGATGCGCTCAACCTCTGCTTCTTGATCAT-----AGCCCTGTGGCTT	2484
QY	3919	GTGACTACTAGAAATTTGCTTTGGTGCTTAAACGGGTCCATCTGTACTGAAAAGGAGG	3978
DB	2485	CAGATATGACACTTTACAGCACTGGGCGCCGCAAAACTCAGGTTCTCTGAAACAAGACAAT	2544
QY	3979	TGCTGACGTGCACTCTTTGGCAAGAAACAGGAGGTGAAATAGAAAAATTAATGCCATGG	4038
DB	2545	TCGTTACATGTATTATTGTCTCAAGAGGAGCAAGAAGTTAAAGTGGAAAGCAGGCAATGG	2604
QY	4039	TATTATCGGCTGTGTCAGAAATCTACTGCTTTAACCCAGCAGACAGGGGAAAACCATAG	4098
DB	2605	TCTTGGCAGCATTTGTTTCAGAGATCAACTGTATTATCAAAAACCAAGATTAATTTATTC	2664
QY	4099	AACTCTCAGGAAAGCCCTAGACCCACTTTTCATGGATCCAGACTTGGCATATGGAACTT	4158
DB	2665	AAGATCCAGAAAAA---TATGATCCATTAATTCATGACCCCTGATCTCTGTGTGGAAACAC	2721
QY	4159	ATACGAAGAGCTGTGGTCATCTAATCAGCAGTGTGCTGGCAGAAGTATTTTGAAGCTG	4218
DB	2722	ACACTAGAGCTGTGGGCACATATGATGATGCCCATTTGTTGGCAAGGATTTTGTATTCG	2781
QY	4219	TACAGCTGAGCTCTCAGCAGCGCATTCATGTTGACCTTTT-----TGACTTGG	4266
DB	2782	TTCAAGCTAAGAAACAGGAAGGCAACAGAGATTTACGCTTACATACAGCATATGATGTAG	2841
QY	4267	AAAGTGGAGAAATATCTTTGGCCTCTTTGCAAAATCTGTGCAATACTGTGATCCCAATTA	4326
DB	2842	AAAAACGGAATATTCCTTTGGCCCTTTTGTGAATGCTTGAGTAATACTGTTATTTC-----	2896
QY	4327	TTCCCTTTGCAACCTCAAAAGATAAACAGTGCAGATGCCTTGTGCTCAACTTTTGA	4386
DB	2897	-CTGCTGCTTCTCCCAAGAAATATTTTAAACACAGGTTAAATTTTTCAGACCAACCAA	2955
QY	4387	CCCTGGCAGCGTGGATACAGACTGTTCTGGCCAGAATATCAGGTTTAAATATAAGACATG	4446
DB	2956	ATCTGACTCAGTGATTAGACAAATATCTCAGCAAAATAAAGCATTTACAGTTTCTTAGGA	3015
QY	4447	CTAAGGAGAAAAACCAATTCCTATTTCTTTTAATCAAGGAATGGGAGATTCCTTTTGG	4506
DB	3016	AAGAAGAAAGTACTCTTAATAATGCCCTCTACAAGAAATTCAGAAAAATGTGGATGAATTAC	3075
QY	4507	AGTTCCATTCATCCTCAGTTTTGGCGTGTGAGTCTTCGATTAATAATTCAATAGCATCA	4566
DB	3076	AGCTCCCTGAAGGGTTCAGGGCTGATTTTCTGCTCAAGATCCCTTATTTCTGAGAGCAATA	3135


```
QY 2799 ATGGATCCCAATAGTCTTCTTACTGTTACTTCTAGAGGTATGAACCTTGCAGAGGCTTTT 2858
Db 1 ATGGATCCCAACAAAGTCTTCTTACTGTTACTTCTAGAGGTATGAACCTTGCAGAGGCTTTT 60
QY 2859 AACAAAGACCATATCTACAAAGACCCAGGATTTGATTAACAATATATACACTAATAGAA 2918
Db 61 AACAAAGACCATATCTACAAAGACCCAGGATTTGATTAACAATATATACACTAATAGAA 120
QY 2919 GAAATGCTTCAGTCCCTCATCTATATTTGTTGGTGAGCGTTATGTACCTGGAGTGGGAAT 2978
Db 121 GAAATGCTTCAGTCCCTCATCTATATTTGTTGGTGAGCGTTATGTACCTGGAGTGGGAAT 180
QY 2979 GTGACCAAGAGAGGTCACAATGAGAGAAATCAATTCACCTTGTCTTTCATTTGAACCCCATG 3038
Db 181 GTGACCAAGAGAGGTCACAATGAGAGAAATCAATTCACCTTGTCTTTCATTTGAACCCCATG 240
QY 3039 CCACACAGTCCCATTTGCCAAAATTTACCTGAGAAATTAATGAACCTGGCTTAGAG 3098
Db 241 CCACACAGTCCCATTTGCCAAAATTTACCTGAGAAATTAATGAACCTGGCTTAGAG 300
QY 3099 AATGTCATAACAAAGTGGCCACATTTAAGAAACCCAGTGTATCAGGCCATGGAGTTTAT 3158
Db 301 AATGTCATAACAAAGTGGCCACATTTAAGAAACCCAGTGTATCAGGCCATGGAGTTTAT 360
QY 3159 GAACCTAAAGATGAATCACTGAAGAGCTTCAATATGTACTTTTATCATTTACTCCAAAACC 3218
Db 361 GAACCTAAAGATGAATCACTGAAGAGCTTCAATATGTACTTTTATCATTTACTCCAAAACC 420
QY 3219 CAGCATAGCAAGGCTGAACATATGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3278
Db 421 CAGCATAGCAAGGCTGAACATATGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
QY 3279 GCATTGCCGCCACACACCTCTGTAATTTGCTGCTTTCAGCAAGTCAATTAACCTT 3338
Db 481 GCATTGCCGCCACACACCTCTGTAATTTGCTGCTTTCAGCAAGTCAATTAACCTT 540
QY 3339 CTCAACTGTGATATCATATGATGATCATCTCAGGACCGTATTTGAGCGGCAATAGACACA 3398
Db 541 CTCAACTGTGATATCATATGATGATCATCTCAGGACCGTATTTGAGCGGCAATAGACACA 600
QY 3399 GATTCTAACTTTGGAGCCGAGGATGCTCCAAATGCTTTTCATATTTCTGGCATTTGGT 3458
Db 601 GATTCTAACTTTGGAGCCGAGGATGCTCCAAATGCTTTTCATATTTCTGGCATTTGGT 660
QY 3459 TTACTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3518
Db 661 TTACTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
QY 3519 TATCATAAAGCTTCAAGATTGGGAAGTTTCAGCCATGAATATACAAATGCTTTTGGAAAAA 3578
Db 721 TATCATAAAGCTTCAAGATTGGGAAGTTTCAGCCATGAATATACAAATGCTTTTGGAAAAA 780
QY 3579 CTCAAAGGAATTTCCCGATTAGAGCCGAGAGGACATGATAAGCTGGATATCTCAGATG 3638
Db 781 CTCAAAGGAATTTCCCGATTAGAGCCGAGAGGACATGATAAGCTGGATATCTCAGATG 840
QY 3639 TTTGACACATGAGCGATTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3698
Db 841 TTTGACACATGAGCGATTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
QY 3699 TCGGAATCTATTAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3758
Db 901 TCGGAATCTATTAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 960
QY 3759 AAAGCTGAAGCTGCTAGGCTACATCGCCAGAGATCATGGC 3799
Db 961 AAAGCTGAAGCTGCTAGGCTCATCGCCAGAGATCATGGC 1001
```

RESULT 9
AAC86934
ID AAC86934 standard; cdna; 1001 BP.

```
XX AAC86934;
AC 02-APR-2001 (first entry)
DT Nucleotide sequence of a partial human Ubql protein.
DE
XX Ubql; E3-type protein; ubiquitin system; ubiquitin-protein ligase;
KW N-end rule pathway; intracellular pathogen; Lysteria monocytogenes;
KW Yersinia enterocolitica; muscle wasting; infection; ss.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT 1..999
FT /*tag= a
FT /*product= "Ubql"
FT /*note= "partial sequence"
XX
XX US6159732-A.
XX
XX 12-DEC-2000.
XX
XX 11-JAN-1999; 99US-0228317.
XX
XX 02-DEC-1997; 97US-0982956.
XX
XX (CALY ) CALIFORNIA INST OF TECHNOLOGY.
XX
XX Kwon YT, Varshavsky A;
XX
XX WPI: 2001-090278/10.
XX P-PSDB; AAB31163.
XX
XX Inhibiting the N-end rule pathway in mammalian cells for treating
XX infections and various diseases associated with muscle tissue wasting,
XX by inhibiting the expression of Ubql gene
XX
XX Claim 4; Column 27-30; 18pp; English.
XX
XX The present sequence encodes a partial Ubql enzyme. Ubql is an E3-type
XX protein of the ubiquitin system. Specifically, it is a ubiquitin-protein
XX ligase. The enzyme is specific for destabilising residues exposed at
XX the N-terminus of protein substrates. Inhibition of the expression of
XX Ubql gene in a cell results in inhibition of the N-end rule pathway.
XX The method is used for treatment of mammalian cells infected with an
XX intracellular pathogen, e.g. Lysteria monocytogenes or Yersinia
XX enterocolitica. Inhibition of N-end rule pathway is also useful for
XX treating various diseases associated with wasting of muscle tissue and
XX infections.
XX
XX Sequence 1001 BP; 363 A; 186 C; 205 G; 247 T; 0 other;
```

```
Query Match 15.8%; Score 996.2; DB 22; Length 1001;
Best Local Similarity 99.7%; Pred. No. 1.1e-261;
Matches 998; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2799 ATGGATCCCAATAGTCTTCTTACTGTTACTTCTAGAGGTATGAACCTTGCAGAGGCTTTT 2858
Db 1 ATGGATCCCAACAAAGTCTTCTTACTGTTACTTCTAGAGGTATGAACCTTGCAGAGGCTTTT 60
QY 2859 AACAAAGACCATATCTACAAAGACCCAGGATTTGATTAACAATATATACACTAATAGAA 2918
Db 61 AACAAAGACCATATCTACAAAGACCCAGGATTTGATTAACAATATATACACTAATAGAA 120
QY 2919 GAAATGCTTCAGTCCCTCATCTATATTTGTTGGTGAGCGTTATGTACCTGGAGTGGGAAT 2978
Db 121 GAAATGCTTCAGTCCCTCATCTATATTTGTTGGTGAGCGTTATGTACCTGGAGTGGGAAT 180
QY 2979 GTGACCAAGAGAGGTCACAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3038
Db 181 GTGACCAAGAGAGGTCACAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
```


Db	488	GCATGGAGAGTGTAAATCGAAGCAGGTTGCCCATTTTCAAGAAACCTGGGATTAACAGGACGAG	547
Qy	3151	GAGTTTATGAACATAAAGATGAATCACTGAAAGACTTCAATATGTACTCTTTATCATTAAT	3210
Db	548	GCATGTATGAACATAAACCAGAAATGTCCAAAGAGTTCAACTTGTATTTCTATCATCACTTTT	607
Qy	3211	CCAAACCCAGCATAGCAAGCGTCAACATATGACGAAGAAAGAGGAAACACAGAAACA	3270
Db	608	CAAGGCGAGACAGTCCAAAGCAGAAAGCGCAACGGAATTTGAAAGACAAATATAG	667
Qy	3271	AAGATGAAGCATTCGCCGCCACCACTCCTGAAATTTCTGCCCTGCTTTCAGCAAAAGTGA	3330
Db	668	AAGATACAGCACTCCCACTCCGGTGTGGCTCCATTTCTGCCCTCTGTTTGGCAAGCCTGG	727
Qy	3331	TTAACTTCTCAACTGTGATATCATGATFACATTTCTCAGAGCCGATATTTGAGCGGCA	3390
Db	728	TTAAATTTTGCAGTCAGATGTCATGTTGTGTCATGTCATGGCAACAATTTCTGCAATGGCTG	787
Qy	3391	TAGACACAGATTTCACTTGTGGACCGAAGGGATGTCCTCAAAATGGGCTTTTTCATATCTGG	3450
Db	788	TGGAACATAATGGATATGCCGTGTCAGAGTCCATGCTGTCGCAAGGGTGTACATTTAATG	847
Qy	3451	CATTGGGTTTACTAGAGAGAACACACAGCTTCAAAAGCTCTCTGAG---AAGAAGTAA	3507
Db	848	GCATGGCCTACAGAAGAAACACCATTTAGAGAAATGTCACGGAAGACATGTATGATA	907
Qy	3508	CATTGGACTTTTATCATPAAGCTTCAAGATTTGGGAAGTTCAGCCATGAAT-----A	3558
Db	908	CATTACCTTCACTCAGAAGATATCAAACTGGTGAAGCGCCAAATAATCTCTTAGCA	967
Qy	3559	TACAAATGCTTTTGGAAAACTCAAGGAATTTCCCACTGTTGAGAGCCAGAAGGACATGA	3618
Db	968	TACTAGCTATGCTGGAAACACTACAAATGCTCCTACCTAGAGTCCACAAAGACATGA	1027
Qy	3619	TAAGTCGATCTACATGTTTGACACAGTGAAGCATTAAGAGAAAAATCTTGTTTAA	3678
Db	1028	TTGGTGGATTTGAAGACTTTTAATGCTGTTAAAAAGATGAGGGAGA-----GTTAC	1081
Qy	3679	TTGTAGCAACACATCAGGATCGGAATCTATTGAAGATGATGAGATTACTCATGATAAG	3738
Db	1082	CTACCACTCCGTGGCAGACAGACAGGAAGAACCATATGGAAGAGATTCAAGGGACAAAG	1141
Qy	3739	AAAAGCAGAACGAAAAAGCTGAAGCTGCTAGGCTACATCGCCAGGAAGATCATGG	3798
Db	1142	ACAAAGCTGAGAGGAAGAAAAAGCAGAGATTGCCAGACTGCGCAGAGAAAGATCATGG	1201
Qy	3799	CTCAGATGCTGCTTACAGAAAAACTTCAATGAACTCATAACTCATGATGACAAATA	3858
Db	1202	CTCAGATGCTGAAATGACGGCATTTTATTGATGAAACAAAGAACTCTTTCAGCAGA	1261
Qy	3859	CATCAGAAATGCTTGGGAAAGAGATTCCATTTATGGAGGAAGAGACACCCAGCAGTCA	3918
Db	1262	CATTAGAACTGGATGCTCAACCTGCTGCTTCTTGATCA-----TAGCCCTGTGGCTT	1315
Qy	3919	GTGACTACTTAGAATTTGCTTTGGGCTTAAAGGGGTCCATCTGTTACTGAAAGGAGG	3978
Db	1316	CAGATATGACATTTACAGCACTGGCCCGCCACAACTCAGGTTCTCGAACAAGACAAT	1375
Qy	3979	TGCTGAGTCGATCCCTTGGCCAGAGACAGAGAGTGAATAATAGAAAATAATGCGCATG	4038
Db	1376	TCGTTACATGATATTGTGTCAAGAGGACGAAGAATTTAAAGTGGAAAGCAGGGCAATGG	1435
Qy	4039	TATTATCGGCTGTGTCAGAAATCTACTGCTTAAACCAGACAGGGGAAACCCATAG	4098
Db	1436	TCITGGCAGCATTTGTTTCAGAGATCAACTGTATTATCAAAAAACAGAAGTAAATTTATTC	1495
Qy	4099	AACCTCTCAGAGAGCCCTAGACCCCACTTTTTCATGATCCAGACTTGGCATATGGAACTT	4158
Db	1496	AAGATCCAGAAAAATAT--GATCCATTTATCATGCAACCTGATCTGTTGTGGAAAC	1552
Qy	4159	ATACAGAAAGCTGTGTCATGTAATGACGACAGTGTGTCGGCAAGAGTATTTTGAAGCTG	4218
Db	1553	ACACTAGTACTGCTGGGCACATTTATGATGCCCATTTGTTGGCAAGGATTTTGTGATTCG	1612

Db 663 CAAGGCGAGAACAGTCCAAAGCGAGAAGAACGCGAAATTTGAAAAGACAAAATAGAG 722
QY 3271 AAGATGAAGCATTTGCCGCCACACCACCTCCTGAAATCTGCCCTGCTTTGAGCAAAAGTGA 3330
Db 723 AAGATGAGACATCCCACCTCCGGTGTGTGCTCCATCTCGCTCTGTTGCAAGCGCTGG 782
QY 3331 TTAACCTCTCAACTGTGATATCATCTCTCAGGACCGTATTTTGGCGGGCAA 3390
Db 783 TTAACATTTTGCAGTCAGATGTCATGTTGTCATCATGGAACAAATCTGCAATGGGCTG 842
QY 3391 TAGACACAGATTTAACTTTGTGGACCGAAGGGATGCTCCAAATGGCTTTTTCATATCTGG 3450
Db 843 TGAACATAATGATATGCTGCTCAGAGTCCATGCTGCAAGGGTGTTCATTAATTTG 902
QY 3451 CATTTGGTCTTACTAGAGAGAACACACAGCTTCAAAAAGCTCCTGAAG---AAGAATAA 3507
Db 903 GCATGCACTACAGAAGAAAACACATTTAGAGAATGTCAGGAGAGCATGTAGTAA 962
QY 3508 CATTTGACATTTTATCAATAGGCTTCAAGATTTGGAAGTTTCAGCCATGAAT-----A 3558
Db 963 CATTTACCTTCACTAGAGATATCAAAACCTGGTGAAGCGCCAAAATAATTCCTTAGCA 1022
QY 3559 TACAAATGCTTTTGGAAAACCTCAAAAGGAATCCCAAGTTAGAGGCGCAGAAAGCATGA 3618
Db 1023 TACTAGCTATGCTGGAACACTACAAATGCTCCTACCTAGAGTCCCAAGACATGA 1082
QY 3619 TAACGTGGATCTAGATGTTTGGACACAGTGAAGCGATTTAAGAGAAAATCTGTGTTAA 3678
Db 1083 TTCGGTGGATTTGAAGACTTTTAAATGCTGTTAAAAAGATGAGGGAGA-----GTTAC 1136
QY 3679 TTGTACCAACCACATCAGATCGGAATCTATTAAAGATGATGAGATTACTCATGATAAG 3738
Db 1137 CTACCAAGTCCCGTGGCAGACAGAGAGAACCAATAATGGAAGAGAGTTCAAGGACAAAG 1196
QY 3739 AAAAAAG 3798
Db 1197 ACNAACTGAG 1256
QY 3799 CTCAGATGCTGCTTACAGAAAACCTTCAATGAACTCAATAACATCATGTATGACATA 3858
Db 1257 CTCAGATGCTGAAATGCAGCGCATTTTATGATGAAAACAAAGAACTCTTTTCAGCAGA 1316
QY 3859 CATCAGAAATGCTGGGAAGAAGATTCATTTATGAGGAGAGAGAGAGAGAGAGAGAGAG 3918
Db 1317 CATTAAGATGGATGCTCAACCTGCTGCTTCTGATCA-----TAGCCCTGTGGCTT 1370
QY 3919 GTGACTACTAGAAATTTGCTGGGTCTTAAAGCGGTCCATCTGTTACTGAAAAGAGG 3978
Db 1371 CAGATATGACACTTACAGCACTGGGCGCCGCAAACTCAGTTCCTGAAACAAGACAAT 1430
QY 3979 TGCTGACGTGCAFTCTTGGCAAGAGAACAGGAGGTGAAAATAGAAAATAATGCCATGG 4038
Db 1431 TCGTTACATGATATTTGTCAAGAGAGAGCAAGAAGTTAAAGTGAAGCAGGCGCATGG 1490
QY 4039 TATTATGGCTGTGTCCAGAAATCTACTGCGCTTAACCCAGCAGAGGGGAAAAACCCATAG 4098
Db 1491 TCTTGGCAGATTTGTTTCAGAGATCAACTGATTTATCAAAAAACAGAAGTAATTTATTC 1550
QY 4099 AACTCTCAGGAGAGCCCTAGACCCACTTTTCATGATCCAGACTTGGCATATGGAACTT 4158
Db 1551 AAGATCCAGAAAATAAT---GATCCATATTATGACCCCTGATCTGTTGTGGAAACAC 1607
QY 4159 ATACAGGAAGCTGTGTCATGTAATGCACGCAAGTGTGCTGGCAGAGAGTATTTTGAAGCTG 4218
Db 1608 ACATGATGCTGTGGCAGCATATGATGCCATGCCATTTGTTGGCAAGGATTTTGTATCCG 1667
QY 4219 TACAGCTGAGCTCTCAGACGCGCATTTCTTGAACCTTTT-----TGACTGG 4266
Db 1668 TTCAACTAAAGACAGCGNAGGACACAGAGATTAAGCTTACATACGAGCTATGATGATAG 1727
QY 4267 AAAGTGGAATAATCTTTGCCCTCTTTTGGAAATCTCTGTGCAATCTGTGATCCCCCATTA 4326
Db 1728 AAAACGAGAAATTCCTTTGCCCTTTGTGAATGCTTGAGTAATGAGTACTGTTATTCCTCTGC 1787

QY 4327 TTCCTTTCCAAACCTCAAAAGATAAACAGTGAAGATGCAGATGCTCTTCTGCTCAACTTTTGA 4386
Db 1788 TGCTTTT-----CTCCAAGAAATATTTTAAACAACAGGTTAAATTTTTCAGACCAACCAA 1841
QY 4387 CCCTGGCAGCGTGGATACAGAGTGTCTGGCCAGAAATATCAGGTTATATATATAAGACATG 4446
Db 1842 ATCTGACATCAGTGGATTAGAACAAATATCTCAGCAATTAAGACATTAACAGATTTCTTAGGA 1901
QY 4447 CTAAAGGAGAAAACCCAAATCCTATTTTCTTAATCAAGGAATGGAGATTTACTTTGG 4506
Db 1902 AAGAAGAAAGTACTCCTAATAATGCTCTACAAGAAATTCAGAAAATGTGGATGAATAC 1961
QY 4507 AGTTCCATTCCTGAGTTTGGCGTTGAGTCTTCGATTAATATATCAAAATAGCATCA 4566
Db 1962 AGCTCCCTGAAGGGTTACGGCTGATTTTTCGCTTAAGATCCCTTATTTCTGAGAGCATAA 2021
QY 4567 AGAAAATGGTTATTCTCTTTGCCACAACAAATTTATAGAATTTGGAATGGAAGTGCACCTG 4626
Db 2022 AAGAAATGCTAACGACATTTGGAATGCTACCTACAGGTGGAGACTAAGGTTTCATCCCA 2081
QY 4627 ATGAAGGGATCTCGAGTCCCATGCTGACCTGGAGCAGCTGCGCTTTCACTATCCAGG 4686
Db 2082 ATGAAGAGGATCTCGTGTCCCAATAATGTTGGGTAGCTGCGCGTACACCATCCAATA 2141
QY 4687 CAATTGAAAATCTATTGGGAGATGAAGAAAACCTCTGTTTGGAGCATCTTCAAAATAGGC 4746
Db 2142 GCATAGAAGAAATTTTGAAGTGAAGATAAACCAATGTTTGGTCTTTTACCTTTCGAGAC 2201
QY 4747 AGCATATGCTCTGAAGCATTAAATGCAAGTTGCAAGTTCGACAGAGGATTTACCTGCTC 4806
Db 2202 TGGATGACTGCTTTAGGTCAITGAGAGATTTGCGCGACACATGGGAGTGGCATCAG 2261
QY 4807 AGGTCTGTATACAGAAACATCTGTTCTCTATCATGTTCTTCTTAACATAAAAAT 4866
Db 2262 TTTCAAGTGGTCCAGGACATTTTCTAAACITTTTGCATCACTGTTGCTTATGACAGCC 2321
QY 4867 CAGAAGATACACCATGCTCTGCTATAGATCTGTTTCAATGTTTGGTGGGTGCTGCTG 4926
Db 2322 ATGAGAACTTCCATGTCATATTAGATAITGACATGTTTCAATTTATTTGGTGGCTTGGTGC 2381
QY 4927 TAGCATTTCCCATCTTGTATTGGGATGACCTGTTGATCTGCAGCCTTCTTCACTAGTT 4986
Db 2382 TTGCATTTCTCGCTTGCAGTG-----TCAGGATTTTTCAGGGAATCAGCC 2426
QY 4987 CTTCCTATAACCACTTTTATCTTCTTCCATTTGATCACCATGGCAGACATGCTTCAGATAC 5046
Db 2427 TTGCACTGGAGACCTTCCACATTTTCCATCTGGTTACTATGCGACACATCATACATCT 2486
QY 5047 TACTTACAGTAGACAC---AGGCCCTTACCCCTTGTCTGAGTTTCAAGAAAGACAGTGAAGG 5103
Db 2487 TACTTACTCTATGTACAGAAGAGAAATGGCATGGATCAAGAAATCCCCCTTGTGAAGAAG 2546
QY 5104 CTCAATCCGATCTTCTTCTTTCAGAGAAATTTCTCAATATACAGTGGCTCCATGGGT 5163
Db 2547 AATCAGCAGTTCTTGTCTTGTATATAAACACTCACACAGTATACGGAAGTGCCTTGAAG 2606
QY 5164 GTGATATCTCGCTGGTGTATTGTTGGGTCTCCTCACTGAAGAAATGGCATACCCCTTATCTTC 5223
Db 2607 AAATACCATCCGGCTGGCATCTGTGGAGAGTGTGAGAGTGAATCATGCTTTCCTCGA 2666
QY 5224 GCTGTGCTGCAATTTTTCCTTTCCTTACTTGGGTAACTCCGCTGAGGAACATGCAATA 5283
Db 2667 AGTGTCTGCTTATTTTTCATTTAAATGGAGTTCCCTTCCCCACCCGACATCAAG 2726
QY 5284 CCAATTCGAGAGAGGAGTACAGTGCATCTGTAGCTATCTATCTTTTACCTACAAAT 5343
Db 2727 T---TCCTGGAACAAGCCATTTTGAACATTTATGTAGCTATCTTTTCCCTACCAACAACC 2783
QY 5344 TGTTCCTGCTCTCCAGGAATATTGGGATACTGTAAAGCCCTTGTCCAGAGGTGGTGTG 5403
Db 2784 TCATTTGCTTTTTCAGAAAATAGTGAGATAATGAATTCATGATTTGAAAGTGGTGGC 2843

Qy 5404 CAGATCCTGCTTACTAACTGTTTGAAGCAAAAACACCGTGGTCAAGTACCCTAGAA 5463
Db 2844 GTACAGTGAAGTTAAAGATATCTAGAGGTGAAGAGATGCTATAGATATCAAGAG 2903
Qy 5464 AAGAAATAGTTGATAGAGCTTCCGTGATGACTATAGCTGCTCCTGAATCAAGCTTCTC 5523
Db 2904 AATCTAACAAAATTAATAACCTCCAGAGGATTACAGCAGCTCATTAATCAAGCATCA 2963
Qy 5524 ATTTGAGTGGCCAGGCTGTCAGATGATGAGCGAAGCATCTGTCCTCTGCTTTTCT 5583
Db 2964 ATTTCTGTCGCCGAAATCAGGTGGTGATAGAGCAGAGCCCAACTCTGTGCTTGTGT 3023
Qy 5584 GTGGGCTATACTATGTTCTCAGAACATTTGCTGCCAGGAATTTGTAACGGGGNAGAG 5643
Db 3024 CGGATCTCTGCTGTGCTCCAGAGTACTCTGCCAGACTGAACCTGGAAGGGGAGGATG 3083
Qy 5644 TTGAGCTTGGATTTTTCACGCACTTCACTGTGGAGCGGAGTCTGCAATTTTCTTAAAAA 5703
Db 3084 TAGGAGCTGCACAGCTCACACCTACTCTGTGCTCTGGAGTGGGCACTTCTCTGAGAG 3143
Qy 5704 TCAGAGAAATCCGAGTGGTCTGTTGAAGTTAAAGGTAAGCCAGAGGTGTCCTATCCAGCTC 5763
Db 3144 TACGGGAATGTCAGGTGCTATTTTAGCTGGCAAAACCAAGGCTGTTTTTATTCTCCTC 3203
Qy 5764 CTTACTTGGATGAATATGGAAACACAGACCCCTGSCCTGAAGAGGGGCAACCCCTTCATT 5823
Db 3204 CTTACCTTGATGACTATGGGAGAGCCGACCGAGGACTCAGACGGGGAATCCTTTACATT 3263
Qy 5824 TATCTGCTGAGCGGTATCGGAAGCTCCATTTGGTCTGGCAACAACACTGCATTTAGAG 5883
Db 3264 TATCCAAGAGCGATTCAAGAGATTTCAGAAGCTCTGGCACCACACACAGTGTCCACAGAG 3323
Qy 5884 AGATTGCTAGGAGCCAGAGACTAATCAGATGTTTATTGGATTCAACTGGCA 5935
Db 3324 AAATTGGACATGCACAGGACCAATCAGACACTGGTTGGCATTTGACTGGCA 3375

RESULT 12
AAA02411
ID AAA02411 standard; cDNA; 733 BP.
XX
AC AAA02411;
XX
DT 19-MAY-2000 (first entry)
XX
DE Human colon cancer cell line polynucleotide sequence SEQ ID NO:2402.
XX
KW Human; colon cancer; tumour; diagnosis; gene expression product;
KW probe; detection; cancerous state; metastasis; identification;
KW breast cancer; oestrogen receptor-positive breast cancer; therapy;
KW oestrogen receptor-negative breast cancer; lung cancer; ss.
XX
OS Homo sapiens.
XX
XX WO9958675-A2.
XX
PD 18-NOV-1999.
XX
XX 13-MAY-1999; 99WO-US10602.
XX
PR 14-MAY-1998; 98US-0085426.
PR 15-MAY-1998; 98US-0085537.
PR 15-MAY-1998; 98US-0085696.
PR 21-OCT-1998; 98US-0105234.
PR 27-OCT-1998; 98US-0105877.
XX
XX (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
PI Lanson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;

XX WPI; 2000-126369/11.
DR
XX Polynucleotide library used to determine cancerous states of mammalian
PT cells -
PT
XX
PS
XX Claim 1; Page 958; 1097pp; English.
XX
CC AAA00010 to AAA02716 represent polynucleotides isolated from cDNA
CC libraries constructed from human colon cancer cell lines. The present
CC invention also describes a method of detecting differentially expressed
CC genes correlated with a cancerous state of a mammalian cell, comprising
CC detecting at least one differentially expressed gene product in a test
CC sample derived from a cell suspected of being cancerous, where detection
CC of the differentially expressed gene product is correlated with a
CC cancerous state of the cell from which the test sample was derived.
CC The polynucleotides sequences can be used in a method for detecting
CC differentially expressed genes correlated with a cancerous state of a
CC mammalian cell. the polynucleotides can also be used as probes for
CC detecting and mapping related genes. They can be used in diagnosis and
CC prognosis of diseases and disorders (e.g. identification of
CC pre-metastatic or metastatic cancerous states, stages of cancer, or
CC responsiveness of cancer to therapy). This is particularly for breast
CC cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
CC negative breast cancer, lung cancer, and colon cancer.
XX
SQ Sequence 733 BP; 218 A; 145 C; 152 G; 210 T; 8 other;
Query Match 10.4%; Score 656.8; DB 21; Length 733;
Best Local Similarity 97.5%; Pred. No. 7.3e-169;
Matches 696; Conservative 0; Mismatches 13; Indels 5; Gaps 3;
Qy 1405 CATATGACCACGCTCATATACAGCCTACAAAGAGCTCTTGACTGTGAGCTGCGAGAGGCC 1464
Db 20 CACGAGACCCTCATATACAGCCTACAAAGAGCTCTTGACTGTGAGCTGCGAGAGGCC 79
Qy 1465 AGTTGATACCACTGCGCATTTGACAAAGAGGGTCTGGGGCTGTTAAAGCGGAGCTATG 1524
Db 80 AGTTGCTNTTCCACTGCCATTTGACAAAGAGGGTCTGGGCTGTTAAAGCGGAGCTATG 139
Qy 1525 CTGCTTGGCAGGAAGCAAGAGAGATATAAAGAGTCAATTCAGAAAAATGCTCTCAACATC 1584
Db 140 CTGCTTGGCAGGAAGCAAGAGAGATTTAAAGAGTCAATTCAGAAAAATGCTCTCAACATC 139
Qy 1585 CACTTCATGTAGAAGTATTACACTCAGAGTATTGCTCATCAGAAATTTGCTTTCGCTC 1644
Db 200 CACTTCATGTAGAAGTATTACACTCAGAGTATTGCTCATCAGAAATTTGCTTTCGCTC 259
Qy 1645 TTGGTTCTGGATGAACAAATATTAGCTATTCAAGTGAAGTCTTAGGAGATCTTTTGGC 1704
Db 260 TTGG-TCTTGGATGAACAAATATTAGCTATTCAAGTGAAGTCTTAGGAGATCTTTTGGC 318
Qy 1705 AAGCATGCTTAGAGAAGAACCTGACTCGAGAGTCCCTCTCTCATTAAGCAGCTTAATGC 1764
Db 319 AAGCATGCTTAGAGAAGAACCTGACTCGAGAGTCCCTCTCTCATTAAGCAGCTTAATGC 378
Qy 1765 TTTGGGATGCAAGCTTTTATAAAGGTGCCCGTAAGATCCCTTCATGAATTTGATCTTCAGCA 1824
Db 379 TTTGGGATGCAAGCTTTTATAAAGGTGCCCGTAAGATCCCTTCATGAATTTGATCTTCAGCA 438
Qy 1825 GTTTTTTTATGGAGATGGAATACAAAAACTCTTTTGTCTATGGAATTTGTGAAGTATTATA 1884
Db 439 GTTTTTTTATGGAGATGGAATACAAAAACTCTTTTGTCTATGGAATTTGTGAAGTATTATA 498
Qy 1885 AACAACTGCAGAAAGAAATATATCAGTGATCATCATCAGAACTATCTCTATTAAGTGCAC 1944
Db 499 AACAACTGCAGAAAGAAATATATCAGTGATCATCATCAGAACTATCTCTATTAAGTGCAC 558
Qy 1945 TTTTCAGTTTCAGATGTTTACTGTTCCCTACTCTGGCTCGACATCTTATTGAAGCAGAAATG 2004
Db 559 TTTTCAGTTTCAGATGTTTACTGTTCCCTACTCTGGCTCGACATCTTATTGAAGCAGAAATG 618
Qy 2005 TTATCTCTCTCATTTACTGAAACTCTGCTAGAAGTTTTTACCTGAGTACTTTGGACAGG-AA 2062

|||||
619 TTATCTCTGTCATTACTGAAACTCTGCTAGAAAGTTTACCTGAGTACTTGACNNNGAAC 678
2063 CAATAAATCAACTTCAGGTTATAGCC--AGCACAAATGGGAAGATATAT 2114
|||||
679 CAATAAATCAACTTCAGGTTATAGCCAGGACAAATTTGGGAAGATATAT 732

RESULT 13
AAH07621
ID AAH07621 standard; cDNA; 712 BP.

AC AAH07621;
26-JUN-2001 (first entry)

DE Human cDNA clone (5'-primer) SEQ ID NO:4456.
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

OS Homo sapiens.
PN EP1074617-A2.

XX 07-FEB-2001.
XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300253.
XX 11-JAN-2000; 2000JP-0118776.
XX 02-MAY-2000; 2000JP-0183767.
XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -

XX Claim 1; SEQ ID 4456; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX Sequence 712 BP; 229 A; 132 C; 156 G; 190 T; 5 other;

Query Match 10.3%; Score 647; DB 22; Length 712;
Best Local Similarity 97.6%; Pred. No. 3.5e-166;
Matches 687; Conservative 0; Mismatches 13; Indels 4; Gaps 3;
QY 1195 AACCTGGAAGACGAGTACTATAAAGAGAATTCACGCTGTCGGTTGAATGAAGAGGTAA 1254
DB 8 ACCTGGGAAGACGAGTACTATAAAGAGAATTCACGCTGTCGGTTGAATGAAGAGGTAA 67
QY 1255 TTGTCGAAGCCAGGAAATATTTCCCTTCAGTGATAAAATATGTCGTAGAAATGACATATAT 1314
DB 68 TTGTCGAAGCCAGGAAATATTTCCCTTCAGTGATAAAATATGTCGTAGAAATGACATATAT 127
QY 1315 GGAAGAGGAAAAAGAACTGCTCTCTGAACTCCAGATAAGGGAGAAAAATGAAGATACT 1374
DB 128 GGAAGAGGAAAAAGAACTGCTCTCTGAACTCCAGATAAGGGAGAAAAATGAAGATACT 187
QY 1375 ATTGTGCTCTTTTCAATGATGAACACCATTCATATGACCACGTCATATACAGCTACAAA 1434
DB 188 ATTGTGCTCTTTTCAATGATGAACACCATTCATATGACCACGTCATATACAGCTACAAA 247
QY 1435 GAGCTCTTGACTGTGAGCTCGCAGAGGCCAGTTGATACCTACCTGTCATGACAAAGAGG 1494
DB 248 GAGCTCTTGACTGTGAGCTCGCAGAGGCCAGTTGATACCTACCTGTCATGACAAAGAGG 307
QY 1495 GTCGTCGGGCTGTTAAAGCGGGAGCTTATGCTGCTTCCAGGAAGCAAAAGGAAGATATA 1554
DB 308 GTCGTCGGGCTGTTAAAGCGGGAGCTTATGCTGCTTCCAGGAAGCAAAAGGAAGATATA 367
QY 1555 AGAGTCATTGAGAAAAATGCTCTCAACATCCACTTCATGTAGAAGTATTACACTCAGAGA 1614
DB 368 AGAGTCATTGAGAAAAATGCTCTCAACATCCACTTCATGTAGAAGTATTACACTCAGAGA 427
QY 1615 TTATGGCTCATCAGAAAAATTTGCTTGGCTTGGTTCCTGGATGAACAAAAATTTATGAGCT 1674
DB 428 TTATGGCTCATCAGAAAAATTTGCTTGGCTTGGTTCCTGGATGAACAAAAATTTATGAGCT 487
QY 1675 ATTCAAGTGACTTTAGGCGAGATCTTTTGCCAGACGCTCCCTAGAGAAGAACCTGACTCGG 1734
DB 488 ATTCAAGTGACTTTAGGCGAGATCTTTTGCCAGACGCTCCCTAGAGAAGAACCTGACTCGG 547
QY 1735 AGAATCCCTGCTCATAGCAGGTTAAATGCTTTGGATGCAAGCTTTTATAAGGTG-CC 1793
DB 548 AGAATCCCTGCTCATAGCAGGTTAAATGCTTTGGATGCAAGCTTTTATAAGGTGCCC 607
QY 1794 CGTAGAGTCCCTTCATGAATTTGATCTTCAGCAGCTTTTATTTTGG-AGATGGAATACAAAA 1852
DB 608 CGTAGAGTCCCTTCATGAATTTGATCTTCAGCAGCTTTTATTTTANGSNGGATGATACAAAA 667
QY 1853 ACTCTTTGCTATGGAATTT--GTGAAGTATTATAACAACACTGCA 1894
DB 668 ACTCTTTGCTATGGAATTTTGTGAAAGTATTATAACAACACTGCA 711

RESULT 14
AAA02327
ID AAA02327 standard; cDNA; 756 BP.
XX AAA02327;
XX AC AAA02327;
XX DT 19-MAY-2000 (first entry)
XX Human colon cancer cell line polynucleotide sequence SEQ ID NO:2318.
KW Human; colon cancer; tumour; diagnosis; gene expression product;
KW probe; detection; cancerous state; metastasis; identification;
KW breast cancer; oestrogen receptor-positive breast cancer; therapy;
KW oestrogen receptor-negative breast cancer; lung cancer; ss.
OS Homo sapiens.
XX XX
XX WO9958675-A2.
XX PD 18-NOV-1999.

XX PF 13-MAY-1999; 99WO-US10602.
XX PR 14-MAY-1998; 98US-0085426.
XX PR 15-MAY-1998; 98US-0085537.
XX PR 15-MAY-1998; 98US-0085696.
XX PR 21-OCT-1998; 98US-0105234.
XX PR 27-OCT-1998; 98US-0105877.
XX (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
PI Lamson G, Drmanac R, Kirkenjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
XX WPI; 2000-126369/11.
XX Polynucleotide library used to determine cancerous states of mammalian
PT cells -
XX PS Claim 1; Page 916-917; 1097pp; English.
XX CC AAA00010 to AAA02716 represent polynucleotides isolated from cDNA
CC libraries constructed from human colon cancer cell lines. The present
CC invention also describes a method of detecting differentially expressed
CC genes correlated with a cancerous state of a mammalian cell, comprising
CC detecting at least one differentially expressed gene product in a test
CC sample derived from a cell suspected of being cancerous, where detection
CC of the differentially expressed gene product is correlated with a
CC cancerous state of the cell from which the test sample was derived.
CC The polynucleotides sequences can be used in a method for detecting
CC differentially expressed genes correlated with a cancerous state of a
CC mammalian cell. The polynucleotides can also be used as probes for
CC detecting and mapping related genes. They can be used in diagnosis and
CC prognosis of diseases and disorders (e.g. identification of
CC pre-metastatic or metastatic cancerous states, stages of cancer, or
CC responsiveness of cancer to therapy). This is particularly for breast
CC cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
CC negative breast cancer, lung cancer, and colon cancer.
XX SQ Sequence 756 BP; 219 A; 145 C; 154 G; 219 T; 19 other;
Query Match 10.1%; Score 638.6; DB 21; Length 756;
Best Local Similarity 95.9%; Pred. No. 7.2e-164;
Matches 677; Conservative 0; Mismatches 26; Indels 3; Gaps 3;
QY 1405 CATATGACACGCTATATACAGCCTACAAAGAGCTCTTTGACTGTGAGCTCGCAGAGGCC 1464
DB 54 CACGAGACCGCTATATACAGCCTACAAAGAGCTCTTTGACTGTGAGCTCGCAGAGGCC 113
QY 1465 AGTTGCATACCACTCCCATTCACAAGAGGTCGCGGGCTTTAAAGCGGAGCTTATG 1524
DB 114 AGTTGCATACCACTCCCATTCACAAGAGGTCGCGGGCTTTAAAGCGGAGCTTATG 173
QY 1525 CTGCTTGGCAGGAAGGAGATATAAAGAGTCATTACAGAAATGCTCTCACATC 1584
DB 174 CTGCTTGGCAGGAAGGAGATATAAAGAGTCATTACAGAAATGCTCTCACATC 233
QY 1585 CACTTCATGTAGAGTATTACACTCAGAGTATTGGCTCATCAGAAATTTGCTTGCCTC 1644
DB 234 CACTTCATGTAGAGTATTACACTCAGAGTATTGGCTCATCAGAAATTTGCTTGCCTC 293
QY 1645 TTGGTTCCTGGATGAACAAATATAGCTATTCAAGTGACCTTTAGGAGATCTTTTGGC 1704
DB 294 TTGGTTCCTGGATGAACAAATATAGCTATTCAAGTGACCTTTAGGAGATCTTTTGGC 353
QY 1705 AAGCATGCTTAGAGAAAGCACTGACTCGGAGAAATCCCTGTCTCATACAGGTTAATGC 1764
DB 354 AAGCATGCTTAGAGAAAGCACTGACTCGGAGAAATCCCTGTCTCATACAGGTTAATGC 413
QY 1765 TTTGGGATGCAAGAGCTTTTATAAAGGTGCCCGCTTAAGATCCTTTCATGAATGTCTTCAGCA 1824

DB 414 TTTGGGATGCAAGAGCTTTATAAAGGTGCCCGTAAAGATCCCTTCATGAATGTCTTCAGCA 473
QY 1825 GTTTTTTTATGGAGATGGAATACAAAAAAGCTCTTTGCTATGGAATTTGTGAAGTATTATA 1884
DB 474 GTTTTTTTATGGAGATGGAATACAAAAAAGCTCTTTGCTATGGAATTTGTGAAGTATTATA 533
QY 1885 AACAACTGCAGAAAGAATATATACAGTATGATCATGACAGAAAGTATCTCTATAACTGCAC 1944
DB 534 AACAACTGCAGAAAGAATATATNAGTATGATCATGACAGAAAGTATCTCTATAACTGCAC 593
QY 1945 TTTCACTTCAGATGTTTACTGTTCTTACTCTGCTGCTGACATCTTATTTGAAGACAGAAATG 2004
DB 594 -TTCACTTCAGATGTTTACTGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 651
QY 2005 TTATCTCTGCTCATTACTGAAACCTCTGCTAGAACCTTTTACCTGAGTACTTGGACAGGAACA 2064
DB 652 TTATCTCTGCTCATTACTGAAACCTCTGCTAGAACCTCTGCTAGAACCTTTTACCTGAGTACTTGGACAGGAACA 711
QY 2065 ATAAATTCACCTCCAGGGTTATAGCCAGGACAAATTTGGGAAGAGT 2110
DB 712 NTAAATTCNACTTCCANGGTTAT-GCCNGGACANATTGGNAAGATT 756
RESULT 15
ID ABZ70161 standard; cDNA; 2616 BP.
XX AC ABZ70161;
XX DT 24-APR-2003 (first entry)
XX DE Fibrinogen 9.57 coding sequence.
XX KW Fibrinogen 9.57; tumour; cytostatic; inflammation; HIV infection;
XX KW immunological disease; haemopathy; anti-HIV; gene; ss.
XX OS Unidentified.
XX FH Key Location/Qualifiers
XX CDS 415..678
XX FT /*tag= a
XX FT /product= "Fibrinogen 9.57"
XX CN13611119-A.
XX PN 31-JUL-2002.
XX PD 26-DEC-2000; 2000CN-0135939.
XX PF 26-DEC-2000; 2000CN-0135939.
XX PR (BODE-) BODE GENE DEV CO LTD SHANGHAI.
XX PA Mao Y, Xie Y;
XX PI WPI; 2002-751538/82.
XX DR P-PSDB; ABP59141.
XX PT New polypeptide fibrinogen 9.57 and polynucleotides encoding this
XX PS polypeptide -
XX PS Claim 6; Page 24-25 (Disclosure); 33pp; Chinese.
XX CC The present sequence is the coding sequence for fibrinogen 9.57. The
XX CC protein is useful for treating various diseases, such as malignant
XX CC tumours, inflammations, immunological diseases, haemopathy and HIV
XX CC infection.
XX SQ Sequence 2616 BP; 780 A; 449 C; 512 G; 875 T; 0 other;
Query Match 9.9%; Score 626.8; DB 24; Length 2616;
Best Local Similarity 99.7%; Pred. No. 2.6e-160;

		Matches	628;	Conservative	0;	Mismatches	2;	Indels	0;	Gaps	0;
QY	5679	GCCGGAGTCTGCATTTTCTAAATAATCAGAGATGCCGAGTGTCTCTGGTTGAAGTAAA									
Db	1	GCCGGAGTCTGCATTTTCTAAATAATCAGAGATGCCGAGTGTCTCTGGTTGAAGTAAA									
QY	5739	GCCAGAGGCTGTGCCCTATCCAGCTCTACTTGGATGAATATGAGAAACAGACCCCTGGC									
Db	61	GCCAGAGGCTGTGCCCTATCCAGCTCTACTTGGATGAATATGAGAAACAGACCCCTGGC									
QY	5799	CTGAGAGGGGCAACCCCTTCATTTATCTCGTGAGCGGTATCGGAAGTCCCATTTGGTC									
Db	121	CTGAGAGGGGCAACCCCTTCATTTATCTCGTGAGCGGTATCGGAAGTCCCATTTGGTC									
QY	5859	TGGCAACACACTGCATTATAGAAGAGATTGCTAGGAGCCAGAGACTAATCAGATGTTA									
Db	181	TGGCAACACACTGCATTATAGAAGAGATTGCTAGGAGCCAGAGACTAATCAGATGTTA									
QY	5919	TTTGGATTCAACTGGCAGTTACTGTGAGCTCCAACCTCGCTCAAGACAATCACAATGA									
Db	241	TTTGGATTCAACTGGCAGTTACTGTGAGCTCCAACCTCGCTCAAGACAATCACAATGG									
QY	5979	CGACAGTAGTAAGGCTGATTCAAAATTTATGGAATTTCTCGAGGCTGGGAAAGTATT									
Db	301	CGACAGTAGTAAGGCTGATTCAAAATTTATGGAATTTCTCGAGGCTGGGAAAGTATT									
QY	6039	GGAGGCTCTTTGCTCCATGTCCAGGTTCACTTACATCAATAAATATTCTTAAATGGAG									
Db	361	GGAGGCTCTTTGCTCCATGTCCAGGTTCACTTACATCAATAAATATTCTTAAATGGAG									
QY	6099	TATTGCTTCAATTAGCAACATATGCTTCACAGGAAAGGACATAGATCAATCTGTT									
Db	421	TATTGCTTCAATTAGCAACATATGCTTCACAGGAAAGGACATAGATCAATCTGTT									
QY	6159	TTATGCTAGTATTTCAGGAATTATTCCCTTCATAATTGTCTCATTTTAT									
Db	481	TTATGCTAGTATTTCAGGAATTATTCCCTTCATAATTGTCTCATTTTAT									
QY	6219	TTCATCCACTTGGTAGATCAAGTCAGTCAACAGTTGTAGACATTTTATGTGTGTTA									
Db	541	TTCATCCACTTGGTAGATCAAGTCAGTCAACAGTTGTAGACATTTTATGTGTGTTA									
QY	6279	ACTCTTCTGCAATTTTGTATTTGGTGT									
Db	601	ACTCTTCTGCAATTTTGTATTTGGTGT									

Search completed: September 27, 2003, 01:57:39
Job time : 1013.89 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 27, 2003, 01:01:23 ; Search time 267.924 Seconds
(without alignments)
10391.910 Million cell updates/sec

Title: US-09-724-126A-1
Perfect score: 6308
Sequence: 1 gccagaattgcacgagg.....aatttgattgtggtttt 6308

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/lna/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/lna/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/lna/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/lna/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/lna/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/lna/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4435.4	70.3	6395	2	US-08-982-956-1
2	4435.4	70.3	6395	3	US-09-228-317-1
3	996.2	15.8	1001	2	US-08-982-956-2
4	996.2	15.8	1001	3	US-09-228-317-2
5	57.4	0.9	7218	1	US-08-232-463-14
6	52.2	0.8	7218	1	US-08-232-463-14
7	45.6	0.7	111282	4	US-09-754-250-3
8	44.4	0.7	7333	3	US-09-138-024-21
9	44.4	0.7	7333	4	US-09-404-066-21
10	44.4	0.7	7333	4	US-09-573-322-21
11	44.2	0.7	72928	3	US-09-009-913-1
12	42.8	0.7	50000	4	US-09-146-053-4
13	42.2	0.7	325	3	US-08-991-789A-236
14	42.2	0.7	325	4	US-09-062-451-236
15	42.2	0.7	325	4	US-09-598-326-236
16	42.2	0.7	325	4	US-09-289-198-236
17	41.2	0.7	36551	4	US-09-738-894A-3
18	41.2	0.7	36551	4	US-09-964-469-3
19	41.2	0.7	42571	4	US-09-810-347-3
20	41.2	0.7	42571	3	US-08-963-602-6
21	40.6	0.6	41708	4	US-09-470-512A-3
22	40.4	0.6	152331	3	US-09-128-155-16
23	40.2	0.6	9510	4	US-09-453-702B-256
24	39.8	0.6	2861	1	US-08-299-953-1
25	39.8	0.6	2861	1	US-08-459-415-1
26	39.8	0.6	2861	4	US-09-066-687-1
27	39.8	0.6	2861	5	PCT-US95-11231-1

c 28	39.8	0.6	3881	1	US-08-299-953-2	Sequence 2, Appli
c 29	39.8	0.6	3881	1	US-08-459-415-2	Sequence 2, Appli
c 30	39.8	0.6	3881	4	US-09-066-687-2	Sequence 2, Appli
c 31	39.8	0.6	3881	5	PCT-US95-11231-2	Sequence 2, Appli
c 32	39.8	0.6	4376	1	US-08-119-125A-1	Sequence 1, Appli
c 33	39.4	0.6	751	4	US-09-687-698-11	Sequence 11, Appli
c 34	39.2	0.6	64467	4	US-09-803-671B-3	Sequence 3, Appli
c 35	39	0.6	32042	4	US-09-245-281-44	Sequence 44, Appli
c 36	39	0.6	32042	4	US-09-340-620A-63	Sequence 63, Appli
c 37	38.8	0.6	246240	2	US-08-724-394A-20	Sequence 20, Appli
c 38	38.8	0.6	246240	2	US-08-724-394A-21	Sequence 21, Appli
c 39	38.8	0.6	246240	2	US-08-724-394A-22	Sequence 22, Appli
c 40	38.4	0.6	5703	4	US-09-280-590A-36	Sequence 36, Appli
c 41	38.4	0.6	18596	3	US-09-318-448-11	Sequence 11, Appli
c 42	38.4	0.6	18597	4	US-09-962-665-8	Sequence 8, Appli
c 43	38	0.6	1785	4	US-09-601-198-156	Sequence 156, App
c 44	37.8	0.6	19011	1	US-08-310-356-36	Sequence 36, Appli
c 45	37.8	0.6	19557	5	PCT-US92-06300-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-982-956-1
; Sequence 1, Application US/08982956
; Patent No. 5861312
; GENERAL INFORMATION:
; APPLICANT: Varshavsky, Alexander
; APPLICANT: Kwon, Yong Tae
; TITLE OF INVENTION: NUCLEIC ACID ENCODING MAMMALIAN UBRI
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kevin M. Farrell, P.C.
; STREET: P.O. Box 999
; CITY: York Harbor
; STATE: ME
; COUNTRY: US
; ZIP: 03911
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/982,956
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Farrell, Kevin M.
; REGISTRATION NUMBER: 35,505
; REFERENCE/DOCKET NUMBER: CIT-2001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (207) 363-0558
; TELEFAX: (207) 363-0528
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6395 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 115..5385
US-08-982-956-1

Query Match 70.3%; Score 4435.4; DB 2; Length 6395;
Best Local Similarity 86.3%; Pred. No. 0;
Matches 4997; Conservative 0; Mismatches 721; Indels 69; Gaps 6;
QY 590 TTCCAGGGCCGTCGTAAGTGTGCTCCTCTC-TCCGACCGCCACAGTTTCGCT 648
|||||

D	b		8	TTCAGGGCCGCTCGTAAAGTGTCGCCCTCGTCGCGTGCGGCCGGCCACAGGTTTCGCGCT	67
Q	y		649	TGCTCTGGCCGGGGGTTCGCAACTCAGCGCTCAGTTTTCCCTCAAGATGGCGACGAGG	708
D	b		68	AGCTGGCGCCGGGGTTCGGAACCTGGGCGTTCGTTTCCCTTAAGATGGCGACGAAG	127
Q	y		709	AGGCTGGAGTGACTGAGAGGATGGAATCAGCGCGAGTTACCCCAGACCCCTCACGCTC	768
D	b		128	AGATTGACGGCGCGAGAGGATGGAGTCAAGCTCAGCCGGAGCCTCCCTTGGCCCCGACGGC	187
Q	y		769	TGCACTCTTGGTGGGATCACAAGTTGATTTTATCTACTGCTTTCCTTTCGATCATTTGGCAC	828
D	b		188	CGGACTCGTGGGATCACAAGTTGATTTCTATACTGCTTCTTCTACATTCATTTGGCAC	247
Q	y		829	AATTGGTGCAGAAAATTTACTTTGCTGAAATGACGCCAGACTTTGGAAAAACGACGAGAAA	888
D	b		248	AATTAGTGCCAGAAAATTTATTTTGTCTGAGATGGACCAGATTTTGGAAAACAGAAGAGA	307
Q	y		889	GTTGTACAAATGTCAAATATTCACCTCCACTGGAAATGGTACTTTATTTGGAGAAGATCCAGATA	948
D	b		308	GTGTACAGATGTCAATACTCACCTCCTTTGGATGGTACTTTATTTGGAGAGGATCCGGATA	367
Q	y		949	TTTTCCTTAGAGAAAATGCAACACACAGTGGAGCATTTTCAGCTTTGTGGGAGGGTTTTCAAAA	1008
D	b		368	TTTTCCTTAGAGAAAATTAACACACAGTGGAGCGTTCCAGTTGTGTGGGAAGGTTTTCAAAA	427
Q	y		1009	GTGGAGAGAACACCTATTCTTTCGAGGATGTGCAATTTGATCCAACATGTGTACTCTGTA	1068
D	b		428	GTGGAGAAACAATATTCCTGTGAGGATTTGCAATTTGATCCAACGTTGTGCTCTGTA	487
Q	y		1069	TGGACTGCTTCCAGGACAGTGTTCATAAAAAATCATCGTTACAAGATGCATCTACTCTG	1128
D	b		488	TGGACTGCTTCCAAAGTAGTGTTCATAAAAAACCATCGTTACAAGATGCATCTACTCTG	547
Q	y		1129	GAGGAGGGTCTGTGACTGTGGAGACACAGAGCATGGAAAACTGGCCCCTTTTTGTGTAA	1188
D	b		548	GAGGGGGCTTCTGTGACTGTGGAGACACAGAGCGTGGAAAACTGGCCCCTTTTTGTGTGG	607
Q	y		1189	ATCATGAACCTGGAAGACAGGTACTATAAAGAGAAATTCACGCTGTCCCTTCAATGAAG	1248
D	b		608	ATCAGAGCCTGGNAGACAGGTACTACAAGAAGAGACTTCATTTGCCCATTCANYGAAAG	667
Q	y		1249	AGGTAATTTGCCAAGCCAGGAAAATATTTCCCTTCAGTGATPAAATATGTGCTAGAAATGA	1308
D	b		668	AGGTGATTTGCTCAAGCCAGGAGAATATTCCTTCTCGGTGATAAAATACATTTGTAGAAATGA	727
Q	y		1309	CTATATGGGAAGAGAAAAGACTGCCCTCTGNACTCCAGATCAAGGGAGAAAAATGAAA	1368
D	b		728	CTATATGGGAAGAAAGAAATGCGCTCTGNACTGCAGATCAAGGGAGAAAAATGAAAC	787
Q	y		1369	GATACTATTGTGCTTTTCAATGATGAACACCATTCATATGACACAGCTCATATACAGCC	1428
D	b		788	GATACTATTGTGCTTTTCAAGATGAGCACCATTCGTATGATCATGTGATCTACAGTC	847
Q	y		1429	TACAAAGAGCTCTTGACTGTGAGCTGCGAGAGGCCAGTTTGCATACCACTGCCATTTGACA	1488
D	b		848	TGCAGAGAGCTTAGATTGCGAGCTTCAGAGGCCACAGCTGCACAGCTGCCATCGACA	907
Q	y		1489	AAGAGGTCTGCGGGCTGTTTAAAGCGGGAGCTTATGCTGCTTGGCAGGAGCAAGGAAG	1548
D	b		908	AAGAGGTCTCGGGCTGTCAAAGCAGAGTGTGATGCCACTTGCAGAGAGCAAGAGG	967
Q	y		1549	ATATAAGAGTTCATTCAGAAAATGCTCTCAACATCCACTTCATGTAGAAGTATTACACT	1608
D	b		968	ATATAAGAGTCACTAGAGAAAGTCTCTCAGCACCCCCCTCCATGTGGAAAGTGTCTGCACT	1027
Q	y		1609	CAGAGATTATGGCTCATCAGAAAATTTGCTTTGGGTCTTGGTTCCTGGATGAACAAAATTA	1668
D	b		1028	CCGTGGTTATGGCTCACAGAAAATTCGCTCTGGGCCCTTGGCTCCTGGATGAACAAAATTA	1087
Q	y		1669	TGACTATTCAAGTGACTTTAGGCAGATCTTTTGCCAAAGCATGCTCTAGAGAGAACCTG	1728
D	b		1088	TGACTATTCAAGTGACTTTAGACAGATATTTTGGCAGGCTGCTCTGTAGAGAGAACCTG	1147

Qy	1729	ACTCGGAGAAATCCCTGTCTCATAAAGCAGGTTAATGCTTTGGGATGCAAAAGCTTTTATAAAG	1788
Db	1148	GCTCTGAAAATCCCTGTCTTATTAAGCAGACTAATGCTTTGGGATGCAAAAGCTTTTATAAAG	1207
Qy	1789	GTGCCCCTAAGATCCCTTCATCAATGATGCTTCACGAGTCTTTTATGGAGATGGAATACA	1848
Db	1208	GTGCCCCTAAGATCCCTTCATCAATGATGCTTTAGTAGTCTTTTATGGAGATGGAATACA	1267
Qy	1849	AAAAACTCTTTGCTATGGAATTTGTGAAGTATTATAAACAACCTGCAGAAAGAATATATCA	1908
Db	1268	AAAACTCTTTGCTATGGAATTTGTGAAGTATTATAAACAACCTGCAGAAAGACTACATCA	1327
Qy	1909	GTGATGATCATGACAGAAAGTATCTCTATAACTGCACHTTCAGTTCAGATGTTTACTGTTTC	1968
Db	1328	CGCAGCACACGAGAAAGCATCTCCATAACCGCCCTGTCGTCGAGATGCTCACCGTCC	1387
Qy	1969	CTACTCTGGCTCGACATCTTATTGAACAGCAGAAATGTTATCTCTGTCTCATTTACTTGAAACTC	2028
Db	1388	CGACTCTTGGCCGGCATCTTATTGAAGCAGAAATGTTATTTCTGTCTCATTTACTTGAAACGC	1447
Qy	2029	TGCTAGAAGTTTTACCTGAGTACTTTGCACAGGAACAATAAATCAACTTCCAGGGTTATA	2088
Db	1448	TGCTAGAAGTTTTACCTGAAATCTTTGGCAGGAACAATAAATCAACTTCCAGGGTTATA	1507
Qy	2089	GCCAGGACAAATTTGGGAAGTAGTATATGCAGTAAATATGTGACCTAAAGTATATCCTGATCA	2148
Db	1508	GCCAGGACAAACTGSGAAGAGTCTACCGAGTTATATGTGACCTAAAGTATATCCTGATTA	1567
Qy	2149	GCAAAACCACAAATYGACAGAAAGATTAGAATGCAGTTCCCTTGAAGGTTTTTCGATCTTT	2208
Db	1568	GCAAGCCCTGTATATGGACAGAACGATTTAAGAGCGCAGTTTCTTGGAAGGTTTTCCGGTCTTT	1627
Qy	2209	TTTTGAAGATCTTACCTGTATGTCAGGGAATGGAAGAAATCCGAAGACAGGTTTGGAAC	2268
Db	1628	TTCTGAAGATCTTACCTGTATGTCAGGGAATGGAAGAAATCAAGAACAAAGTTTGGACAC	1687
Qy	2269	ACATTGAAGTGGAATCTGATTTGGGAGCTGCCATTTGCTATACAGATGCAATTTGAAGAATA	2328
Db	1688	ACATTGAAGTGGAATCTGACTGGGAGCTGCCATCGCTATACAGATGCAACTAAAGAATA	1747
Qy	2329	TTTTACTCATGTTCCAGAGTGTTGCTGTGTGATGAAGAACATCTTACTTTGTCTGGCTTATA	2388
Db	1748	TTTTGTCTCATGTTCCAGAGTGTTGCTTTGTGATGAAGATCTCTTACTGCTGGCTTATA	1807
Qy	2389	AAGAATGTCAAAAGCTGTGATGAGTGCAGTACCAGTTTCATATCTAGTAGCAAGACAG	2448
Db	1808	AAGAATGTCAAAAGCTGTAAAGTGTGAGTACAAATTTCTCAGTACCAGACAG	1867
Qy	2449	TAGTACAATCGTGTGGACATAGTTTGGAAACAAAGTCCCTACAGAGTATCTGAGGATCTTG	2508
Db	1868	TAGTCCAATTTGCGGGTCATAGTCTGGAACCAATCCTACAAGTGTCTGAGGACCTTG	1927
Qy	2509	TAAGCATACATCTGCCACTCTCTAGGACCCCTTGCTGTCTTCATGTACGTTTAAAGCAGC	2568
Db	1928	TAAGCATACACTGCCACTCTCTAGAACACTTGTGTGCTCTTCATGTACGTTTAAAGCAGAC	1987
Qy	2569	TGGTGCTGTTTCAAGACTGCATCAATTTGTCTCTTTTGGAGGACTTTCAAGTAGAGGTAC	2628
Db	1988	TAGTGTCTATTTCAAGACTGCATCAATTTGTGCTCTTTTGACGCTTTTCAAGTAGAGGTCC	2047
Qy	2629	TAGTGGAAATATCCCTTACGTTGTCTGTTGTTGCTGCCAGGTTGTTGCTGAGATGTGGC	2688
Db	2048	TGTTGGAGTACCCCTGGCTGGCTGGCTGGCTGAGTGTGCTGAGATGTGGC	2107
Qy	2689	GAAGAAATGGACTCTCTCTTATTAGCCAGGTTGTTTTATTACCAAGATGTTTAAGTGCAGAG	2748
Db	2108	GAAGAAAGGGCTCTCACTCATCAGCCAGGTTTTTCTATTATCAAGATGTTTAAATGCAGGG	2167
Qy	2749	AAGAAATGTATGATAAAGATATCATGCTTTCAGATTTGGTGCATCTTTAATGGATCCCCA	2808
Db	2168	AGGAAATGTAGATAAAGATATCATGCTTTCAGATTTGGAGATCTTAAATGGATCCCCA	2227

[illegible]

Db 4388 CCGTGGATCTGCAGCCGTCGCCACTTAGTTCTTTCATATAACCACCTCTATCTCTTCATC 4447

Qy 5017 TGATCACCATTGGCACACATGCTTCAGATATCTACTTACAGTAGACA-----CAG 5064

Db 4448 TGATCACCATTGGGCACATGCTTCAGATATCTCTTACACAGATACACAGATCTGTCTCCAG 4507

Qy 5065 GCCTACCCCTTGCTCAGGTTCAAGACACAGTGAAGAGCTCATTCGCCATCTCTCTTCT 5124

Db 4508 GCGCGCGCTTGCTAGGGTGGAAGAGATAGTGAAGAGGCTCGCTGTCATCTCTCTTCT 4567

Qy 5125 TTGCAGAAATTTCTCAATATACAAAGTGGCTCCATTGGTGTGTGATTTCTGGCTGGGTATT 5184

Db 4568 TTGTGGAAGTGTCCGACACACAGACGGCTCTACTGGGTGGGTGCTCCCGGCTGGTACC 4627

Qy 5185 TGTGGGFTCTCACTGAAGAATGGCATCACCCCTTATCTTCGCTGTGCTGATTTGTTTTC 5244

Db 4628 TGTGGCTCTCCCTGAGGAACGGCATCACCCCTTACCTCCGCTGTGCTGCACTGCTTTTC 4687

Qy 5245 ACTATTTACTTTGGGTAACTCCGCTGAGGAACCTGCATACAAATTCGCGAAGGAGT 5304

Db 4688 ACTATTTACTTTGGGTAACTCCGCTGAGGAACCTGTTTGGCAATTCCTGCGAAGGAAAT 4747

Qy 5305 ACAGTCACCTCTGTAGCTATCTATCTTACCTACAAATTTGTTCCCTCTCTCCAGGAAT 5364

Db 4748 TCAGTCACCTCTGTAGCTATCTATCTTACCCACAAATTTGTTCCCTCTCTCCAGGAAT 4807

Qy 5365 ATTGGGATCTGTAAGCCCTTGCTCCAGAGGTGGTGTGAGATCTCTGCTTACTAACT 5424

Db 4808 ATTGGGATACCAATGAAGCCCTTACTACAGAGGTGGTGTGAGATCTCTGCTTACTCAAGT 4867

Qy 5425 GTTTGAAGCAAAAACACCGTGTGAGTACCCCTAGAAAAGAAATAGTTTGTATAGAGC 5484

Db 4868 CTTTGAAGCAAAAAGTGTGTGCTAGTACCCCTAGAAAAGAAATAGTTTGTATAGAGC 4927

Qy 5485 TTCTGTAGTACTATAGTGTCTCTGAATCAAGCTTCTCATTTTCAGGTGCCAGGCTG 5544

Db 4928 TTCTGTAGGACTACAGTGTCTCTAAATCAGGCTTCTCACTTTAGGTGTCCACGGTCTG 4987

Qy 5545 CAGATGATGAGCAAAAGCATCTCTCTGCTCTGCTTTCTGTGGGCTATACTATGTCTC 5604

Db 4988 CAGATGATGAGCAAAAGCATCTCTCTGCTCTGCTTTCTGTGGGCTATCTGTGTTCTC 5047

Qy 5605 AGAACATTTGCTGCCAGGAAATTTGTGAACGGGGAGAGGTTGGAGCTTGCATTTTTCACG 5664

Db 5048 AGAACATCTCTGCCAAGAAATAGTGAATGGGAAGAGGTTGGAGCGTGGCTTTTCATG 5107

Qy 5665 CACTTCACTGTGGAGCGGAGTCTGCATTTTCTTAAATCAGAGATGCCAGTGGTCC 5724

Db 5108 CGTTCATTTGTGTGGAGTCTGCATTTTCTTAAATCCGAGAAATGCAGGTTGGTCC 5167

Qy 5725 TGGTTGAAGTAAAGCCAGAGGCTGTGCCTTATCCAGCTCTCTTACTTGGATGAATATGGAG 5784

Db 5168 TGGTTGAAGTAAAGCCAGAGGCTGTGCCTTATCCAGCCCTTACTTGGATGAATATGGAG 5227

Qy 5785 AAACAGACCCTGGCTGAAGAGGGGCAACCCCTTCTATTTATCTCGTGAAGCGGTATCGGA 5844

Db 5228 AAACAGACCAGGCTAAAGAGAGGAAACCCACTTCTATTTATCTCGGAGCGGTATCGGA 5287

Qy 5845 AGCTCCATTTGGTCTGCGACACACATGCAATATAGAAGAGATTTGCTAGAGCCCAAGAGA 5904

Db 5288 AGCTGCAATTTGGTCTGCGACACAGCACTGCAATATAGAAGAGATTTGCTAGAGCCCAAGAGA 5347

Qy 5905 CTAATCAGATGTTATTTGGATTCAACTGGCAGTGTACTGTGAGTCCCAACTCTGCCTCAAG 5964

Db 5348 CTAATCAGATGCTATTTGGATTCAACTGGCAGTGTACTGTGAGTCTGAGTCTGCTCAAG 5407

Qy 5965 ACAATCACAATACGACAGTAGTAAAGGCTGATTCAAAAATATGGAAGAACTTTCTGAGG 6024

Db 5408 ACAATCATGATGACATCAATAAAGACTGATCTAAATTTCTAGAGAACTTTCTGAGG 5467

Qy 6025 GCTGGGAAAGTATTTGGAGGCTTTTGTCTCCATGTCCAGGTTCACTTACATCAATAAAT 6084

Db 5468 ACGGGGAGAGTATTTGGAGGCTTTTGTATCCATGTCTCCAGATTTACACACATTAATAAAT 5527

Qy 6085 ATTCTTTAATGGAGTATTCCTTTCAATTTAGCAAAACATATGCTTCACAGGAAAAA-AGGAC 6143

Db 5528 ATTCTTTAATGGAAATATTCCTTTCAATTTAGCAAAACATATGCTTCAGGGGAAAAACAAGAC 5587

Qy 6144 ATAGATCAATCTGT-----TCTT 6160

Db 5588 ATAGATTAATGTTTATTTCTTAGAACACTAAAGAAATGCTTGTTCATCCAAAGTGTCTAT 5647

Qy 6161 ATGCTAGTATTTCCAGAAATTTATTCCTCTTCAATAATTTGCTCATTTTCTATTTTATTT 6220

Db 5648 TTCTGCTAATTTTCCAGAAACTCTTTCCTTTCATACTCTCTAGTTCATTTCTAT 5707

Qy 6221 CATCCACTTGTGTAGTGAAGTCAAGTCAACAGTTGTAGACATTTATGTTGGTTAAC 6280

Db 5708 CACCCACCTGTTAATGAGTCACATTAAGCATTTTGTGGACATTTCTCCATCTGGCTAAC 5767

Qy 6281 TCTTCTGCAATTTGTTATTTGGTGT 6307

Db 5768 ATCTGTGCACCTTTGTATTTGGTGT 5794

RESULT 2

US-09-228-317-1

; Sequence 1, Application US/09228317

; Patent No. 6159732

; GENERAL INFORMATION:

; APPLICANT: Varshavsky, Alexander

; APPLICANT: Kwon, Yong Tae

; TITLE OF INVENTION: NUCLEIC ACID ENCODING MAMMALIAN UBR1

; NUMBER OF SEQUENCES: 2

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Kevin M. Farrell, P.C.

; STREET: P.O. Box 999

; CITY: York Harbor

; STATE: ME

; COUNTRY: US

; ZIP: 03911

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09228,317

; FILING DATE:

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Farrell, Kevin M.

; REGISTRATION NUMBER: 35,505

; REFERENCE/DOCKET NUMBER: CIT-2001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (207) 363-0558

; TELEFAX: (207) 363-0528

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 6395 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 115..5385

US-09-228-317-1

Query Match 70.3%; Score 4435.4; DB 3; Length 6395;

Best Local Similarity 86.3%; Pred. No. 0;

Matches 4997; Conservative 0; Mismatches 721; Indels 69; Gaps 6;

Qy 590 TTCAGGGCCGTCGTAAAGTGTCTGCTGCTC-TCCGACCGGCCACAGGTTTCCGCT 648

Db 8 TTCAGGGCCGTCGTAAAGTGTCTGCTGCTGCGGCCGCGCCACAGGTTTCCGCT 67

QY	649	TGCTCTGCCGGGGTGGCACTGCAGGCGTCAGTTCCTTCCATCAAGATGGCGGAGG	708
Db	68	AGCTGGCGCGGGGTGCGGAACCTCGGGCGTTCGTTTCCCTTAAAGATGGCGGAGG	127
QY	709	AGGCTGGAGGTACTGAGAGGATGGAATCAGCGCGAGTTACCCAGAGCCCTCAGCGTC	768
Db	128	AGATGGACGGCGCGAGAGGATGGAGCTCAGCCGAGCCTCCCTGGCCCGCAGCGGC	187
QY	769	TGGCATCTTGGTGGGATCAGCAAGTTGATTTTATATCTGCTTCTTTCATCATTTGGCAC	828
Db	188	CGGCATCGTGGTGGGATCAGCAAGTTGATTTCTATCTGCTTCTTACATCATTTGSCAC	247
QY	829	AATTGTCGCCAGAAATTTACTTGTCTGTAATGCACCCAGACTTGGAAAGCAGAGGAAA	888
Db	248	AATTAGTGCAGAAATTTTGTGTGAGATGGACCCAGATTTGGAAAAGCAAGAGAGA	307
QY	889	GTGTACAAATGTCAATATTACTCCACTCGAATGGTACTTATTGTGAGAGATCCAGATA	948
Db	308	GTGTACAGATGTCAATACTACTCTTGTGAGTGGTACTTATTGTGAGGATCCGATA	367
QY	949	TTTGCTTAGAGAAATTTGAACACAGTGGAGCAATTCAGCTTTGTGGGAGGTTTCAAAA	1008
Db	368	TTTGCTTAGAGAAATTTAAAACACAGTGGAGCGTTCCAGTTGTGTGGGAGGTTTCAAAA	427
QY	1009	GTGGAGAGACAACCTATTCTTGCAGGGATGTGCAATTTGATCCAAATGTGTACTCTGTA	1068
Db	428	GTGGAGAAACAATATTCTCTGTAGGATTTGCAATTTGATCCAACTGTGTCTCTGTA	487
QY	1069	TGGACTGTCTCAGGACAGTGTTCATAAAATCATCGTTACAGATGTCATACTCTACTG	1128
Db	488	TGGACTGTCTCAGGACAGTGTTCATAAAATCATCGTTACAGATGTCATACTCTACTG	547
QY	1129	GAGGAGGTTCTGTGACTGTGGAGACACAGAGGCGATGGAAACTGGCCCTTTTGTGTAA	1188
Db	548	GAGGAGGTTCTGTGACTGTGGAGACACAGAGGCGTGGAAACTGGCCCTTTTGTGTG	607
QY	1189	ATCATGAACCTGGAAGACAGGTACTATAAAAGAGAAATTCACGCTGTCCGTTGAAATGA	1248
Db	608	ATCAGAGCCTGGAAGACAGGTACTACAAAAGAGAGCTTACATTTGCCATTTGAATGA	667
QY	1249	AGGTAAATTTCAAGCCAGGAAATTTCTTCACTGATGATAAATATGCTGATAGATGA	1308
Db	668	AGGTAAATTTCAAGCCAGGAAATTTCTTCACTGATGATAAATATGCTGATAGATGA	727
QY	1309	CTATATGGGAAGAGAAAGAACTGCCCTCCTGAACCTCCAGATAAGGGAGAAAATGAAA	1368
Db	728	CTATATGGGAAGAGAAAGAAATTCCTCCTGAACTGCAGATTAAGGGAGAAAATGAA	787
QY	1369	GATACATTTGTCTCTTTCAACGATGAGACCAATTCGTATGATGATGATGATGATGATG	1428
Db	788	GATACATTTGTCTCTTTCAACGATGAGACCAATTCGTATGATGATGATGATGATGATG	847
QY	1429	TACAAGAGCTCTTGAATGAGTGCAGAGAGGCGGATTTGATACACCTGATGATGATG	1488
Db	848	TGCAGAGAGCTCTAGATTTGCGAGCTTGCAGAGAGACAGCTGACAGAGCTGATGATG	907
QY	1489	AGAGGGTCTGCGGCTGTAAAGCGGAGCTTATCTGCTGCGAGGAGCAAGGAG	1548
Db	908	AGAGGGTCTGCGGCTGTAAAGCGGAGCTTATCTGCTGCGAGGAGCAAGGAG	967
QY	1549	ATATAAGAGTCAATTCAGAAATGTCTTCAACATCCACTTCATGTAGAGTATTACAT	1608
Db	968	ATATAAGAGTCACTCAGAGAGCTCTCTCAGCAGCCCTCCATGTGGAAGTCTGACT	1027
QY	1609	CAGAGATTATGCTCATCAGAAATTTGCTTTGGCTTGTGCTTCTGATGAGCAAAATTA	1668
Db	1028	CGTGGTTATGCTCACCAGAAATTCGCTCTGCGCTTGGCTTCTGATGAGCAAAATTA	1087
QY	1669	TGAGCTATTCAAGTACTTTAGGAGATCTTTTGGCAAGCATGCTTTAGAGAGACCTG	1728
Db	1088	TGAGCTATTCAAGTACTTTAGGAGATCTTTTGGCAAGCATGCTTTAGAGAGACCTG	1147
QY	1729	ACTCGGAGAAATCCCTGTCTCATAGCAGGTTAATGCTTTGGTGTGAAAGCTTTATAAG	1788
Db	1148	GCTCTGAAAATCCCTGTCTTATAAGCAGAGCTAATGCTTTGGATGCAAACTTTATAAG	1207
QY	1789	GTGCCCCAAGATCCTTCATGAATTTGATCTTTCAGCAGTCTTTTATGAGATGGAATACA	1848
Db	1208	GTGCCCCAAGATCCTTCATGAATTTGATCTTTCAGTGTCTTTTATGAGATGGAATACA	1267
QY	1849	AAAACTCTTTGCTTATGGAATTTGAGATTTATAAACAACCTGCGAGAAAGATATATCA	1908
Db	1268	AAAACTCTTTGCTTATGGAATTTGAGATTTATAAACAACCTGCGAGAAAGATATATCA	1327
QY	1909	GTGATGATCATGACAGAGATCTCTAATCTGACCTTTCAGTCTGAGTCTTACTGCTTC	1968
Db	1328	CGCAGCACCAGAGAGAGATCTCTAATCCGCGCTTCCGTCGAGATGCTCACCGTCC	1387
QY	1969	CTACTCTGGCTCGACATCTTATTGAAGACAGAGATTTATCTCTCTCTATTTACTGAACTC	2028
Db	1388	CGACTTTGGCCCGCATCTTATTGAAGACAGAGATTTATTTCTGTCTATTACTGAAACGC	1447
QY	2029	TGCTAGAAGTTTACCTGAGTACTTGGACAGAGACAATAAATTCAACTTCAGGGTTATA	2088
Db	1448	TGCTAGAAGTTTACCTGAGTACTTGGACAGAGACAATAAATTCAACTTCAGGGTTATA	1507
QY	2089	GCCAGGACAAATTCGGAGAGCTATATCCAGTAATATGTGACCTAAGATATATCCGATCA	2148
Db	1508	GCCAGGACAAATTCGGAGAGCTATATCCAGTAATATGTGACCTAAGATATATCCGATCA	1567
QY	2149	GCAAAACCACAATATGGACAGAAAGTTAAGAATGCAGTCTCTTCAAGGTTTTCGATCTT	2208
Db	1568	GCAAGCTGTCTATGACAGACAGATTTAAGAGCGCAGTCTCTGGAAGTTTCCGGTCTT	1627
QY	2209	TTTTGAAGATTTTACCTGTATGCGAGGAATGGAAGAAATCCGAAGACAGCTTTGGCGAAC	2268
Db	1628	TTCTGAAGATTTTACCTGTATGCGAGGAATGGAAGAAATCAGAAGACAAGTTGGACAAC	1687
QY	2269	ACATTGAAGTGGATCTCTGATTGGAGGCTGCCATTTGCTATCAGATGCAATTTAGAATA	2328
Db	1688	ACATTGAAGTGGAGCTTACTGCTGGAGGCTGCCATTTGCTATCAGATGCAATTTAGAATA	1747
QY	2329	TTTTTACATCTTCCAAAGTGTGTCTTGTGATGAAGAACTCTTACTTGTGGCTTATA	2388
Db	1748	TTTTTGTCTATTTCCAAAGTGTGTCTTGTGATGAAGATCTCTTACTTGTGGCTTATA	1807
QY	2389	AAGAATGTCAAAAAGCTGTGATGAGGTGCAGTACCAGTTTTCATATCTAGTACGAAGAC	2448
Db	1808	AAGAATGTCAAAAAGCTGTGATGAGGTGCAGTACAAATTTTCATGCTCCAGTACGAAGAC	1867
QY	2449	TAGTACAACTCGTGGACATAGTTTGGAAACAAAGTCTCAGAGATATCTGAGGATCTTG	2508
Db	1868	TAGTGAATTTGTGGGTCTATAGTCTGGAACCAATCTCAGAAAGTCTGAGGACCTTG	1927
QY	2509	TAAGCATACATCTGCCACTCTTAGGACCTTGTGGTCTTCATCTAGCTTTTAAACGAGC	2568
Db	1928	TAAGCATACATCTGCCACTCTTAGAACACTTGTGGTCTTCATCTAGCTTTTAAACGAGAC	1987
QY	2569	TGGTGTCTGTTTCAAGACTGCATGAATTTGTCTTTTGGAGACTTTTCAAGTAGAGGTAC	2628
Db	1988	TAGTGTCTATTTCAAGACTGCATGAATTTGTGCTTTTGGAGACTTTTCAAGTAGAGGTCC	2047
QY	2629	TAGTGAATATCTTTTACGTTGTCTGTTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG	2688
Db	2048	TGGTGGAGTACCCCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	2107
QY	2689	GAGAAATGAGTGTCTCTTATTAGCCAGGTGTTTTTATTACAGATGTTAAGTGCAGAG	2748
Db	2108	GAGAAACGGGCTCTCACTCATCAGCCAGGTTTTTCTATTATCAAGATGTTAAATCGAGGG	2167
QY	2749	AAGAAATGATGATAAAGATATCATGCTTCAGATTTGCTGCTTCTTAAATGATCCCA	2808
Db	2168	AGAAATGATGATAAAGATATCATGCTTCAGATTTGCTGCTTCTTAAATGATCCCA	2227
QY	2809	ATAAGTTCTTCTTACTTGGTACTTACAGAGGTATGAATTTGCCGAGGCTTTTAAACAAGACCA	2868

Db	2228	ACAAGTCTCTGTTACTGGTACTTTCAGAGATATGAACCTTACTGTACTGTTTAAACAAGACCA	2287
Qy	2869	TATCTACAAAAGACCCAGGATTTGATTAACAACAATAATACACACTAATAGAGAANAATGCTTC	2928
Db	2288	TATCCACAAAAGACCCAGGATTTGATTAACACAGATATATACATTAATAGAGAANAATGCTTC	2347
Qy	2929	AGTCTCTCATCTATATTGTGGGTGAGCGTTATGTACTCGAGTGGGAAATGTGACCAAG	2988
Db	2348	AGGTCTCTCATCTATATTGTGGGAGACGTTATGTACTCGAGTGGGAAATGTTACCAGAG	2407
Qy	2989	AAGAGTCAACAATGAGAGAAATCATCTACTGCTTTTGGCATTTGAACCCATGCCACACAGTG	3048
Db	2408	AGGAGGTTAATAGAGAGATTACTACACTTACTTTGCAATTTGAGCGCATGCCACACAGTG	2467
Qy	3049	CCATTGCGCAAAATTTACCTGAGAAACAATAATGAACCTGCTTAGAGAAATGTCATAA	3108
Db	2468	CCATCGCGCAAAACCTTACCTGAGAAGCAAAATTAATGAACCTGCTTAGAGAAATGTCATAA	2527
Qy	3109	ACAAAGTGGCCACATTTAAGAAACACAGTGATACAGGCCATGGAGTTTATGAACATAAAG	3168
Db	2528	ACAAAGTGGCCACATTTAAGAAACACAGTGATGTCGGGCCATGGAGTTTATGAATGAAG	2587
Qy	3169	ATGAATCACCTGAAGACTTCAATATGTACTTTTATCATTTACTCCAAACCCACAGATAGCA	3228
Db	2588	ATGAATCACCTGAAGACTTCAATATGTACTTTTACCATTATTTCAAAACACAGCATAGCA	2647
Qy	3229	AGGCTGAACATATGACAGAAAGAGAGAAACAAGAAACAAGATGAAGCATTGCCCG	3288
Db	2648	AGGCTGAACATATGACAGAAAGAGAGAAACAAGAAACAAGATGAAGCATTGCCCG	2707
Qy	3289	CACCACACCTCTCGAATTCGCGCTTTCAGCAAAAGTGATTAACCTTCTCAACTGTG	3348
Db	2708	CGCCACCTCTCCAGAGTTCGCGCTTTCAGCAAAAGTAGTCAACCTGCTCAGCTGTG	2767
Qy	3349	ATATCATGATGTACATTTCTCAGGACCGTATTTGAGCGGGCAATAGACAGATTTCAACT	3408
Db	2768	ATGTTATCATATACATCCTCAGGACCATCTTTGAGCGGGCAGTGGACACAGGACTCAATC	2827
Qy	3409	TGTGGACGGAAGGATGCTCCAAATGCTTTTTCATATCTTGGCATTTGGGTTTACTAGAAG	3468
Db	2828	TGTGGACAGAAAGGATGCTCAGATGCGTTCATATATATTGGCACTGGGCTTGTGGAAG	2887
Qy	3469	AGAAAGCAACAGCTTCAAAAAGCTCCTGAAGAAGTAACATTTGACTTTTATCATAAAG	3528
Db	2888	AGAAGCAGCAGCTTCAGAAAGCTCCTGAGAGGAAGTGGCTTTTGACTTTTACCATAAAG	2947
Qy	3529	CTTCAAGATTGGGAAGTTCAGCCAT-----GAATATACAANAATGCTTTTGGAAAAAC	3579
Db	2948	CTTCAAGATTGGGAAGTTCAGCCATGAATGCTCAGAATATACAANAATGCTCTTGGAAAGAC	3007
Qy	3580	TCAAAGGAATTCGCCAGTTAGAAGCCAGAGGACATGATAACGTGGATACTTCAGATGT	3639
Db	3008	TCAAAGGAATTCGCCAATTAAGAAGCCAGAGGACATGATAACATGGATACTCCAGATGT	3067
Qy	3640	TTGACACAGTGAAGCGATTAAAGAAAAATCTTCTTTAAATTGTAGCAACACACATCAGGAT	3699
Db	3068	TTGACACAGTGAAGCGATTAAAGAAAAATCTTCTTTAAATTGTAGTGGCAACACATTCAGAC	3127
Qy	3700	CGGAATCTATTAGAANAATGATGAGATTACTCATGTATGAAGAAAAAGCAGAAAGAA	3759
Db	3128	TGGAGTGCATTAAAGTGAAGGATTTACTCATGTATAAAGAAAGGACGGAAGAA	3187
Qy	3760	AAGCTGAAGCTGTAGGCTACATGCCAGAGATCATGGCTCAGATGTCTGCCCTTACAGA	3819
Db	3188	AAGCTGAAGCGCGTATAGGCTTATCGCCAGAGATTCATGGCCAGATGTCTGCCCTTACAGA	3247
Qy	3820	AAAACTTCATTGAACCTCATAAACTCATGTATGACAATACATCAGAAATGCTCGGGAAG	3879
Db	3248	AAAACTTCATTGAACCCCAAACTCATGTATGATAATACGTCAAGATTAACAGGGAAG	3307
Qy	3880	AAGATTCATTTATGGAGGAAGAGACACCCACAGCAGTCAGTGAATCTCTAGAAATTCGTT	3939
Db			
Db	3308	AGACTCCATTTATGGAGGAAGAGACCACTCTAGCAGTCTAGTGAGGCCTCTAGAATTTGCTC	3367
Qy	3940	TGGGTCTTAAAGGGGTCCATCTGTTACTGAAAAGAGGTGCTGACGTGCATCTCTTTGCC	3999
Db	3368	TGGGCCCTTAACGGGCCCGGCTGTTTACCGAAAAGAGGTGCTGACGTGCATCTCTGCC	3427
Qy	4000	AGAAGAACAAGGAGTGAATAATAGAAAATAAATGCCATGTTATGTCAGCATGTGTGCAGA	4059
Db	3428	AGAAGAACAAGAGTGAATAATAGAAAATAAATGCCATGTTATGTCAGCATGTGTGCAGA	3487
Qy	4060	AATCTACTGCCCTTAAACCAGCACAGGGGAAAACCCATAGAACTCTCAGGAGAGCCCTAG	4119
Db	3488	AATCCACCCCTTAAACCAGCACAGAGGGAAGCCTGTGACCACCTTATAGGGAACACTGG	3547
Qy	4120	ACCACCTTTTCATGGATCCAGACTTGGCATATGGAACCTTATACAGGAAGCTGTGGTCAATG	4179
Db	3548	ACCCTCTTTTCATGGATCCAGACTTGGCACATGGAACCTTATACAGGAAGCTGTGGTCAATG	3607
Qy	4180	TATGACACAGTGTCTGTGCAGAACTATTTTGAAGCTGTACAGCTGAGCTCTCTCAGCAGC	4239
Db	3608	TATGACATGAGTGTCTGTGCAGAACTATTTTGAAGCTGTGAGCTGAGCTGCAGCAGC	3667
Qy	4240	GCATTCATGTTGACCTTTTGTGCTGGAAGTGGAGAAATATCTTTGCCCTCTTTGCAAT	4299
Db	3668	GCATTCATGAGTGTGTTGACCTGGAGGGCGGAGTACCTATATCCCGCTCTGCAAGT	3727
Qy	4300	CTCTGTGCAATACGTGTGATCCCATTTATCTTTGCAACTTCAAAAGATAAACAAGTGAGA	4359
Db	3728	CTCTGTGCAACACTGTATCCCATCCATCCCTTTGACGCGCAGAGATCAACAGTGAGA	3787
Qy	4360	ATGACAGATGCTTTGCTCAACTTTTGACCTTGCACGGTGGATPACAGACTGTTCTGGCCA	4419
Db	3788	ATGCGGAGGCTTTGCTCAACTTTTGACCTTGCAGCTGGATACAGACTGTCTTTGCCA	3847
Qy	4420	GAATATCAGGTTATATATAAGACATGCTTAAAGGAGAAAACC---CAATTTCTATTTTCT	4476
Db	3848	GAATATCGGGTTATATATAAAGCATGCTTAAAGGAGAAAGCCAGCAGTTCCTGTCTGT	3907
Qy	4477	TTAATCAAGGAATGGGAGATTCTACTTTGGAGTTCCATTTCCATCTCCTGAGTTTGGCGTTG	4536
Db	3908	TTAATCAAGGAATGGGAGATTCAACTTTTGAGTTTCACTTCCATCTGAGTTTGGAGTTTC	3967
Qy	4537	AGTCTTCGATTAATAATTCAAAATAGCATCAAGGAAATGTTATTTCTTTTGGCACAAACA	4596
Db	3968	AGTCTTCGSGTAAATATTCAATAGTATCAAGGAAATGGTCACTCTCTTCGCGCACAAACA	4027
Qy	4597	TTTATAGAATTGGAATGAAAGTCCACCTGATGAAAGGATTCCTCAGTCCCCCATCTGA	4656
Db	4028	TTTACAGAATTTGGCCTGAAAGTGCCTCTGATGAACCTAGACCCACGAGTGCCCATGATGA	4087
Qy	4657	CTTGGAGCAGCTCGCTTTCACATATCAGGCAATTTGAAATCTATTGGGAGATGAAGGAA	4716
Db	4088	CTTGGAGCAGCTGTGCGTTCCACCATCAGGCAATCGAAACCTTGTGGGAGATGAAGGAA	4147
Qy	4717	AACTCTCTTTTGGAGCAGCTTCAAAATAGGAGCATTAATGGTCTCTGAAAGCAATTAAT	


```
QY 2859 AACAGACCATCTTACAAAAGACAGGATTTGATTAAACAATATAATACACTAATAGAA 2918
Db 61 AACAGACCATATCTACAAAAGACAGGATTTGATTAAACAATATAATACACTAATAGAA 120
QY 2919 GAAATGCTTCAGTCTCATCTATATTTGTTGGGTGAGCGTTATGTACCTGGAGTGGGAAAT 2978
Db 121 GAAATGCTTCAGTCTCATCTATATTTGTTGGGTGAGCGTTATGTACCTGGAGTGGGAAAT 180
QY 2979 GTGACCAAGAGAGGTCAATGAGAGAAATCATTCACCTTGGTTTGCATTTGAACCCATG 3038
Db 181 GTGACCAAGAGAGGTCAATGAGAGAAATCATTCACCTTGGTTTGCATTTGAACCCATG 240
QY 3039 CCACACAGTCCATTCGCCAAAATTTACCTGAGATGAAAATTAATGAAACTGGCTTAGAG 3098
Db 241 CCACACAGTCCATTCGCCAAAATTTACCTGAGATGAAAATTAATGAAACTGGCTTAGAG 300
QY 3099 AATGTCATAAACAAGTGGCCACATTTAAGAAACCCAGGTGTATCAGGCCATGGAGTTTAT 3158
Db 301 AATGTCATAAACAAGTGGCCACATTTAAGAAACCCAGGTGTATCAGGCCATGGAGTTTAT 360
QY 3159 GAACTAAAAGATGAATCACTGAAAGACTTCAATATGTACTTTTATCATTACTCCAAAACC 3218
Db 361 GAACTAAAAGATGAATCACTGAAAGACTTCAATATGTACTTTTATCATTACTCCAAAACC 420
QY 3219 CAGCATAGCAAGCTCAACATATGTCAGAGAAAGAGAGAAACAAGAAACAAGATGAA 3278
Db 421 CAGCATAGCAAGCTCAACATATGTCAGAGAAAGAGAGAAACAAGAAACAAGATGAA 480
QY 3279 GCATTCGCCACACACCTCCTGAAATCTGCGCTTTCAGCAAGGTGATTAACCTT 3338
Db 481 GCATTCGCCACACACCTCCTGAAATCTGCGCTTTCAGCAAGGTGATTAACCTT 540
QY 3339 CTCACATGATATCATGTATCTCAGGACCGTATTTGAGCGGGCAATAGACACA 3398
Db 541 CTCACATGATATCATGTATCTCAGGACCGTATTTGAGCGGGCAATAGACACA 600
QY 3399 GATTCCTAAGTGTGGACCGAGGATGCTCCAAATGGCTTTTCATATCTGCGCATTTGGT 3458
Db 601 GATTCCTAAGTGTGGACCGAGGATGCTCCAAATGGCTTTTCATATCTGCGCATTTGGT 660
QY 3459 TTACTAGAGAGAGCAACAGCTTCAAAAAGCTCTGAAAGCTCCTGAAGAAAGTAACTTTG 3518
Db 661 TTACTAGAGAGAGCAACAGCTTCAAAAAGCTCTGAAAGCTCCTGAAGAAAGTAACTTTG 720
QY 3519 TATCATAAGGCTTCAAGATTGGGAAGTTACAGCCATGAATACAAATGCTTTTGGAAAAA 3578
Db 721 TATCATAAGGCTTCAAGATTGGGAAGTTACAGCCATGAATACAAATGCTTTTGGAAAAA 780
QY 3579 CTCAAAAGGAATCCCCAGTTAGAGGCCAGAGGACATGATAACGTGGATCTCAGATG 3638
Db 781 CTCAAAAGGAATCCCCAGTTAGAGGCCAGAGGACATGATAACGTGGATCTCAGATG 840
QY 3639 TTTGACACAGTGAAGCGATTAGAGAAAAATCTGTTTAAATTTAGCAACCATCATGGA 3698
Db 841 TTTGACACAGTGAAGCGATTAGAGAAAAATCTGTTTAAATTTAGCAACCATCATGGA 900
QY 3699 TCGGAATCTATTAAAGATGATGATTTACTTCATGATAAAGAAAAAGCAGAACGAAAGA 3758
Db 901 TCGGAATCTATTAAAGATGATGATTTACTTCATGATAAAGAAAAAGCAGAACGAAAGA 960
QY 3759 AAAGCTGAAGCTGCTAGGCTACATCGCCAGAAGATCATGCC 3799
Db 961 AAAGCTGAAGCTGCTAGGCTTTCATCGCCAGAAGATCATGCC 1001
```

RESULT 4

US-09-228-317-2

; Sequence 2, Application US/09228317

; Patent No. 6159732

; GENERAL INFORMATION:

; APPLICANT: Varshavsky, Alexander

; APPLICANT: Kwon, Yong Tae

; TITLE OF INVENTION: NUCLEIC ACID ENCODING MAMMALIAN UBRI

```
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kevin M. Farrell, P.C.
; STREET: P.O. Box 999
; CITY: York Harbor
; STATE: ME
; COUNTRY: US
; ZIP: 03911
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/228,317
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Farrell, Kevin M.
; REGISTRATION NUMBER: 35,505
; REFERENCE/DOCKET NUMBER: CIT-2001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (207) 363-0558
; TELEFAX: (207) 363-0528
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1001 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..999
; US-09-228-317-2
```

Query Match 15.8%; Score 996.2; DB 3; Length 1001;

Best Local Similarity 99.7%; Pred. No. 5.3e-279;

Matches 998; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
QY 2799 ATGGATCCCAATAAGTTCTTGTACTTGTACTTTCAGAGGTATGAACCTTCCGAGGCTTTT 2858
Db 1 ATGGATCCCAACAAGTTCTTGTACTTGTACTTTCAGAGGTATGAACCTTCCGAGGCTTTT 60
QY 2859 AACAGACCATATCTACAAAAGACAGGATTTGATTAAACAATATAATACACTAATAGAA 2918
Db 61 AACAGACCATATCTACAAAAGACAGGATTTGATTAAACAATATAATACACTAATAGAA 120
QY 2919 GAAATGCTTCAGTCTCATCTATATTTGTTGGGTGAGCGTTATGTACCTGGAGTGGGAAAT 2978
Db 121 GAAATGCTTCAGTCTCATCTATATTTGTTGGGTGAGCGTTATGTACCTGGAGTGGGAAAT 180
QY 2979 GTGACCAAGAGAGGTCAATGAGAGAAATCATTCACCTTGGTTTGCATTTGAACCCATG 3038
Db 181 GTGACCAAGAGAGGTCAATGAGAGAAATCATTCACCTTGGTTTGCATTTGAACCCATG 240
QY 3039 CCACACAGTCCATTCGCCAAAATTTACCTGAGATGAAAATTAATGAAACTGGCTTAGAG 3098
Db 241 CCACACAGTCCATTCGCCAAAATTTACCTGAGATGAAAATTAATGAAACTGGCTTAGAG 300
QY 3099 AATGTCATAAACAAGTGGCCACATTTAAGAAACCCAGGTGTATCAGGCCATGGAGTTTAT 3158
Db 301 AATGTCATAAACAAGTGGCCACATTTAAGAAACCCAGGTGTATCAGGCCATGGAGTTTAT 360
QY 3159 GAACTAAAAGATGAATCACTGAAAGACTTCAATATGTACTTTTATCATTACTCCAAAACC 3218
Db 361 GAACTAAAAGATGAATCACTGAAAGACTTCAATATGTACTTTTATCATTACTCCAAAACC 420
QY 3219 CAGCATAGCAAGCTGAAACATATGTCAGAGAAAGAGAGAAACAAGAAACAAGATGAA 3278
Db 421 CAGCATAGCAAGCTGAAACATATGTCAGAGAAAGAGAGAAACAAGAAACAAGATGAA 480
QY 3279 GCATTCGCCACACACCTCCTGAAATCTGCGCTTTCAGCAAGGTGATTAACCTT 3338
```


; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)883-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; CLONE: ptzpt-fls
; IMMEDIATE SOURCE:
;
US-08-232-463-14

Query Match
Best Local Similarity 10.1%; Score 52.2; DB 1; Length 7218;
Matches 48; Conservative 218; Mismatches 211; Indels 0; Gaps 0;

QY 4698 CTATTGGGAGATGAAGAAACCTCTGTTGGAGCACTTCAAAATAGGCAGCATAATGGT 4757
DB 971 CCATCAGTGGCTACTATACTATTTTCCTTCGTTGCCATACGCTCACAGAAATAA 1030
QY 4758 CTGAAGACATTAATGCAGTTGCGAGTTGCACAGAGGATACCTGCTCAGGCTCGTGATA 4817
DB 1031 TTCCGAGCTGGCTGCAGGTCGAGGAGCTTGCATGATGTTGCTGCTGCTGCTGCTGCTG 1090
QY 4818 CAGAAACATCTGGTCTGCTCTTCTATCAGTTGTTCTTCTTCAATACATAAATCAGAGATACA 4877
DB 1091 YY 1150
QY 4878 CCATGCTCTGCTATAGATCTGTTTCACTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 4937
DB 1151 YY 1210
QY 4938 TCCTTGTATGGATGACCTGTGATCTGCAGCTTCTCAGTGTAGTCTTCTCTATAAC 4997
DB 1211 YY 1270
QY 4998 CACCTTATCTCTTCATTTGATCACCATGGCAGCATGCTTCAGATACCTTACAGTA 5057
DB 1271 YY 1330
QY 5058 GACACAGGCTACCCCTGCTCAGTTCAAGAGACAGTGAAGAGGCTCATTCGCAATCT 5117
DB 1331 YY 1390
QY 5118 TCTTTCTTGCAGAAATTTCTCAATATACAAGTGGCTCCATGCTGGTGTGATATTCCT 5174
DB 1391 YY 1447

RESULT 7
US-09-754-250-3
; Sequence 3, Application US/09754250
; Patent No. 6376225
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui et al
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHODIESTERASE
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND USES THEREOF
; FILE OF INVENTION: PHOSPHODIESTERASE PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL001063
; CURRENT APPLICATION NUMBER: US/09/754,250

; CURRENT FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 111282
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(111282)
; OTHER INFORMATION: n = A,T,C or G
US-09-754-250-3

Query Match
Best Local Similarity 0.7%; Score 45.6; DB 4; Length 111282;
Matches 57; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 293 TGATACACACCTGGAAACACCATCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 352
DB 59380 TGCTCCAGCTCCTGGCACCACCATCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 59439
QY 353 GCTGGATCCCGAGCTT 368
DB 59440 GATACCTCAAGTAATT 59455

RESULT 8
US-09-138-024-21
; Sequence 21, Application US/09138024A
; Patent No. 6004779
; GENERAL INFORMATION:
; APPLICANT: Bradley, John D.
; APPLICANT: Thompson, Craig M.
; APPLICANT: Moore, Jeffrey B.
; APPLICANT: Wobbe, C. Richard
; APPLICANT: Healy, Judith M.
; APPLICANT: Donnelly, Caroline E.
; TITLE OF INVENTION: REGULATED GENE EXPRESSION IN YEAST
; FILE REFERENCE: 0342/ID469US1
; CURRENT APPLICATION NUMBER: US/09/138,024A
; CURRENT FILING DATE: 1998-08-21
; EARLIER APPLICATION NUMBER: 60/056,719
; EARLIER FILING DATE: 1997-08-22
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 21
; LENGTH: 7333
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Plasmid pZM197
US-09-138-024-21

Query Match
Best Local Similarity 0.7%; Score 44.4; DB 3; Length 7333;
Matches 81; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 989 TTGTGGAGGGGTTTCAAAAGTGGAGACACACCTATTCTTTCAGGGGATTGTGCAATTCA 1048
DB 3222 TTGTGGAGGGAATTCAAATAGGGGAACCCCTTGTATAGGTCTCATGAGTGTGGTGGCA 3281
QY 1049 TCCAACTGTGTACTCTGTATGAGTCTTCCAGGACAGTGTTCATAAAAAATCATCGTTA 1108
DB 3282 TGATCTGTGTCTGTTTGTATTTTCAATCCAAAGATCATGTGAATCATCATCT 3341
QY 1109 CAAGATGCATCTTCTACTGGA 1130
DB 3342 TTGTACCGATATATCTACTGAA 3363

RESULT 9
US-09-404-066-21
; Sequence 21, Application US/09404066


```
; Sequence 4, Application US/09146053A
; Patent No. 6399349
; GENERAL INFORMATION:
; APPLICANT: Ryan, James W.
; APPLICANT: Sprinkle, Terry Joe Curtis
; APPLICANT: Venema, Richard C.
; TITLE OF INVENTION: Human Aminopeptidase P Gene
; FILE REFERENCE: MCG103
; CURRENT APPLICATION NUMBER: US/09/146,053A
; CURRENT FILING DATE: 1998-09-02
; EARLIER APPLICATION NUMBER: 60/057,854
; EARLIER FILING DATE: 1997-09-02
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 4
; LENGTH: 50000
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-146-053-4

Query Match      0.7%; Score 42.8; DB 4; Length 50000;
Best Local Similarity 87.0%; Pred. No. 0.7;
Matches 47; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 299 CACACCCCTGGAAACCAACCATCTACTTCTGTCTATGAATTTGACTACTCTA 352
DB 23136 CAGCCCTGGCAACCAACCATCTACTTCTGTCTATGAATTTGACTACTCTA 23189

RESULT 13
US-08-991-789A-236/c
; Sequence 236, Application US/08991789A
; Patent No. 6225054
; GENERAL INFORMATION:
; APPLICANT: Frudakis, Tony N.
; APPLICANT: Smith, John M.
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF BREAST CANCER
; NUMBER OF SEQUENCES: 292
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed IP Law Group
; STREET: 701 Fifth Avenue, Suite 6300
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/991,789A
; FILING DATE: 11-Dec-1997
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Potter, Jane E. R.
; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 210121.419C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 236:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 325 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; SEQUENCE DESCRIPTION: SEQ ID NO: 236:

; Sequence 4, Application US/09146053A
; Patent No. 6399349
; GENERAL INFORMATION:
; APPLICANT: Ryan, James W.
; APPLICANT: Sprinkle, Terry Joe Curtis
; APPLICANT: Venema, Richard C.
; TITLE OF INVENTION: Human Aminopeptidase P Gene
; FILE REFERENCE: MCG103
; CURRENT APPLICATION NUMBER: US/09/146,053A
; CURRENT FILING DATE: 1998-09-02
; EARLIER APPLICATION NUMBER: 60/057,854
; EARLIER FILING DATE: 1997-09-02
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 4
; LENGTH: 50000
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-146-053-4

Query Match      0.7%; Score 42.8; DB 4; Length 50000;
Best Local Similarity 87.0%; Pred. No. 0.7;
Matches 47; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 299 CACACCCCTGGAAACCAACCATCTACTTCTGTCTATGAATTTGACTACTCTA 352
DB 23136 CAGCCCTGGCAACCAACCATCTACTTCTGTCTATGAATTTGACTACTCTA 23189
```

```
US-08-991-789A-236

Query Match      0.7%; Score 42.2; DB 3; Length 325;
Best Local Similarity 79.4%; Pred. No. 0.041;
Matches 50; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 298 ACACACCCCTGGAAACCAACCATCTACTTCTGTCTATGAATTTGACTACTCTAGCTGG 357
DB 283 AAAAGCCCTGGCAACCAACCATCTACTTCTGTCTATGAATTTGACTACTCTAGCTGG 224
QY 358 ATC 360
DB 223 CTC 221

RESULT 14
US-09-062-451-236/c
; Sequence 236, Application US/09062451
; Patent No. 6344550
; GENERAL INFORMATION:
; APPLICANT: Frudakis, Tony N.
; APPLICANT: Smith, John M.
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF BREAST CANCER
; NUMBER OF SEQUENCES: 297
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/062,451
; FILING DATE: 04-APR-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.419C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 236:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 325 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
US-09-062-451-236

Query Match      0.7%; Score 42.2; DB 4; Length 325;
Best Local Similarity 79.4%; Pred. No. 0.041;
Matches 50; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 298 ACACACCCCTGGAAACCAACCATCTACTTCTGTCTATGAATTTGACTACTCTAGCTGG 357
DB 283 AAAAGCCCTGGCAACCAACCATCTACTTCTGTCTATGAATTTGACTACTCTAGCTGG 224
QY 358 ATC 360
DB 223 CTC 221
```


RESULT 15

US-09-598-326-236/c
; Sequence 236, Application US/09598326
; Patent No. 6423496
; GENERAL INFORMATION:
; APPLICANT: Fridakis, Tony N.
; Smith, John M.
; Reed, Steven G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF BREAST CANCER
; NUMBER OF SEQUENCES: 247
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed Intellectual Property Law Group PLLC
; STREET: 701 Fifth Avenue, Suite 6300
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/598,326
; FILING DATE: 20-Jun-2000
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Potter, Jane E.R.
; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 210121.419D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 236:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 325 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; SEQUENCE DESCRIPTION: SEQ ID NO: 236:
US-09-598-326-236

Query Match 0.7%; Score 42.2; DB 4; Length 325;
Best Local Similarity 79.4%; Pred. No. 0.041;
Matches 50; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
QY 298 ACACACCCCTGGAAACACCACTTCTACTTCTGTGTATGAATTTGACTACTCTAGCTGG 357
DB 283 AAAGCCCTGGCAACACCACTTCTATTGTGTCTTAATTTGACTACTCTAGCTAC 224
QY 358 ATC 360
DB 223 CTC 221

Search completed: September 27, 2003, 13:39:30
Job time : 277.924 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 27, 2003, 09:29:56 ; Search time 986.772 Seconds
(without alignments)
15914.219 Million cell updates/sec

Title: US-09-724-126a-1
Perfect score: 6308
Sequence: 1 gccagaattcgccagagg.....aattttgtatttggtgtttt 6308

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1678620 seqs, 1244745471 residues

Total number of hits satisfying chosen parameters: 3357240

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_NA.*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	813.6	12.9	3327	10	US-09-529-063-56 Sequence 56, Appl
2	813.6	12.9	3327	12	US-10-414-378-56 Sequence 56, Appl
3	813.6	12.9	3502	10	US-09-529-063-57 Sequence 57, Appl
4	813.6	12.9	3502	12	US-10-414-378-57 Sequence 57, Appl
5	438.4	6.9	505	11	US-09-918-995-36721 Sequence 532, App
6	250.6	4.0	972	9	US-09-822-849A-532 Sequence 7, Appli
c 7	219.8	3.5	4573	13	US-10-071-766-7 Sequence 260533,
c 8	212.2	3.4	578	13	US-10-027-632-260533 Sequence 490, App
c 9	105.6	1.7	1461	9	US-09-925-301-490 Sequence 14952, A
c 10	89.6	1.4	326	10	US-09-960-352-14952 Sequence 14952, A
11	65	1.0	465	11	US-09-918-995-27470 Sequence 8907, Ap
12	60	1.0	60	12	US-09-908-975-8907 Sequence 8115, Ap
13	53.8	0.9	611	13	US-10-027-632-8115 Sequence 204690,
14	49.6	0.8	624	13	US-10-027-632-204690 Sequence 143140,
15	49.4	0.8	781	13	US-10-027-632-143140 Sequence 143138,
16	49.4	0.8	804	13	US-10-027-632-143138

ALIGNMENTS

RESULT 1

US-09-529-063-56
; Sequence 56, Application US/09529063
; Patent No. US20020102542A1
; GENERAL INFORMATION:
; APPLICANT: FUKUSHIMA, DAIKICHI
; APPLICANT: SHIBAYAMA, SHIRO
; APPLICANT: TADA, HIDEAKI
; TITLE OF INVENTION: POLYPEPTIDE, CDNA ENCODING THE POLYPEPTIDE, AND USE OF
; FILE OF INVENTION: THE BOTH
; FILE REFERENCE: Q58769
; CURRENT APPLICATION NUMBER: US/09/529.063
; CURRENT FILING DATE: 2000-04-07
; PRIOR FILING DATE: 1998-10-06
; PRIOR FILING DATE: 1997-10-07
; PRIOR FILING DATE: 1997-10-07
; NUMBER OF SEQ ID NOS: 117
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 56
; LENGTH: 3327
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-529-063-56

Query Match 12.9%; Score 813.6; DB 10; Length 3327;
Best Local Similarity 55.4%; Pred. No. 7.9e-215;
Matches 1856; Conservative 0; Mismatches 1409; Indels 87; Gaps 11;
QY 2632 TCGAATATCTTTACGTTGCTGTGTGTTGTTGCCAGGTGTTGCTGAGATGTGGGAA 2691
DB 8 TAGAACACCTCTTAGATGCTTGTGTGTGCCAAGTACATGCGGAATGTGGAGAA 67
QY 2692 GAATGACGTGCTCTTATTAGCCAGGTGTTTATACCAAGATGTTAAAGTCGACAGAA 2751
DB 68 GAATGGTCTCTCTAGTAACACGATTTATTACTACCAATAGTGAATGCAGACGTG 127
QY 2752 AAATGATGATAAAGATCATCATGCTTCAGATTTGTTGATCTTTAATGATCCCAATA 2811

Db 128 AGATGTTTACAGAGGATGTAGTAATGCTTCAGACAGGTGTCTCCATGATGGATCCAAATC 187
QY 2812 AGTCTTGTGTACTGGTACTCAGAGGTATGAATTCGCCGAGGCTTTTAAACAAGACCATAT 2871
Db 188 ATTTCCTGATGATCATGCTCAGCGCGTTTGAACCTTTATCAGATTTTTCAGTACTCCAGACT 247
QY 2872 CTACAAAA-----GACCAGGATTTGATTAAACAATATAATACAC 2910
Db 248 ATGAAAAAGATTTAGTTCTGAGATTACCCATAGGATGTTGTTCAGCAGAAACAATACTC 307
QY 2911 TAATAGAGAAATGCTTCAGGCTCATCTATATTTGGTGGTGAGCGTTATGTACTCTGGAG 2970
Db 308 TAATAGAGAAATGCTATACCTCATTTAATGCTTTGTTGGAGAGAGATTTAGTCCCTGGAG 367
QY 2971 TGGAAATGTGACCAAGAGAGGTCACATGAGAGAAATCATTCACCTTCTGCTTTGCAATG 3030
Db 368 TTGGACAGGTAAATGCTACAGATGAATCAAGCAGAGATTTATCCATCAGTTGAGTATCA 427
QY 3031 AACCCATGCCACAGTGCCATTGGCCAAAATTTTACTTGAGATGAATAAATGAATAACTG 3090
Db 428 AGCCTATGGCTCATAGTGAATTTGGTAAAGTCTTTTACCTGAAGATGAGAAACAAGAGACTG 487
QY 3091 GCTTAGAGATGTCATAAACAAGTGGCCACATTTAAGAAACAGGTTATTCAGGCCATG 3150
Db 488 GCATGGAGAGTGTATCGAAGCAGTTGCCCCATTTCAAGAAACCTGGATTACAGGACGAG 547
QY 3151 GAGTTTATGAATTAAGATGAATCACTGAAAGACTTCAATATATGTAATTTTATCATTAAT 3210
Db 548 GCATGTATGAATGAACACAGATGTGCCAAAGATTCACACTTGTATTTCTATCACATTT 607
QY 3211 CAAAACCCAGCATAGCAGGCTGAACATATGAGAAAGAAAGAGAGAAACAAGAAACA 3270
Db 608 CAAGGCGAGAACAGTCCCAAGGCAAGAGAAAGCGCAAAATTTGAAAGACAAATAAGAG 667
QY 3271 AAGATGAAGCATTTGCCGCCACCACCTCCTGAAATTCGCCCTGCTTCAGCAAAATGA 3330
Db 668 AAGATACAGACTCCCAACCTCCGGTGTGGCTCCATTTCCGCCCTCTGTTTGCAGACCTGG 727
QY 3331 TTAACCTTCTCAACTGTGTATCATGATGATACATTTCTCAGGACCGTATTTGAGCGGCA 3390
Db 728 TTAACATTTTGCAGTCAGATGTCATGTTGTGCATCATGGCAACAATTTCTGCAATGGGCTG 787
QY 3391 TAGACACAGATTTCTAATGTTGGACCGAAGGATGCTCCAAATGGCTTTTCAATATCTGG 3450
Db 788 TGGAAACATAATGGATATGCTGTGTCAGAGTCCATGCTGCAAAAGGTTGTACATTTAATG 847
QY 3451 CATTTGGGTTTACTAGAGAGAGCAACAGCTTCAAAAAGCTCCTGGAAG---AACAAGTAA 3507
Db 848 GCATGGGCATACAGAAGAAACAAACATTTAGAGAAATGTCACGGAAGAGCATGTAGTAA 907
QY 3508 CATTTGACTTTTATCATAGGCTTCAAGATTGGGAAGTTGAGCCATGAAT-----A 3558
Db 908 CATTTACTTTCACAGAAATATCAAAACCTGGTGAAGGCCCAAAAATTTCTCCTAGCA 967
QY 3559 TACAAATGCTTTTGGAAAACTCAAGGAATTTCCCAAGTTAGAGGCCAGAGACATGA 3618
Db 968 TACTAGCTATGCTGGAACACTACAAAATGCTCCTTACCTAGAAAGTCCACAAAGACATGA 1027
QY 3619 TAAGCTGGATACTACATGTTTGACACAGTGAAGCGATTAAGAGAAAAATCTTGTTTAA 3678
Db 1028 TTCGGTGGATTTGAAGACTTTTAAATGCTGTAAAAAGATGAGGGAGA-----GTTTCA 1081
QY 3679 TTGTAGCAACCATCATCAGATCGGAATCTTAAAGATGATGAGATTACTCATGATMAAG 3738
Db 1082 CTACCACTCCGTTGGCAGACAGCAAGAACCATTAATGGAGAGAGTTCAAGGGACAAAG 1141
QY 3739 AAAAGCAGAACGAAAAAGAAAGCTGAAGCTGCTAGGCTACATTCGCCAGAAAGATCATGG 3798
Db 1142 ACAAGCTGAGAGGAAGAAAAAGCAGAGATTGCGACACTGCGCAGAGAAAAAGATCATGG 1201
QY 3799 CTCAGATGCTGCCCTTACAGAAAAACTCATTTGAACATCAATAACTCATGATACACATA 3858
Db 1202 CTCAGATGCTGTAATCGAGGGCATTTTATTTAGTGAACAAACAAGAACTCTTTTCAGCAGA 1261
QY 3859 CATCAGAAATGCCCTGGGAAAGAGATTCATATTGGAGGAGAGAGACCCACAGCAGTCA 3918
Db 1262 CATTAGAACTGGATGCCCTCAACCTCTGCTGTCTTTGATCA-----TAGCCCTGTGGCTT 1315
QY 3919 GTGACTACTCTAGAATTCCTTTGGTCTCTAAACGGGGTCCATCTGTTACTGAAAGAGGG 3978
Db 1316 CAGATATGACACTTTACAGCACTGGGCCCCGCAACAATCAGGTTCTCTGAACAAAGACAA 1375
QY 3979 TGCCTGAGCTGCATCCTTTGCCAAGAAAGAGGAGGTGAAATAGAAAATATATGCATGG 4038
Db 1376 TCGTTACATGTTATTTGTTCAAGAGGAGCAAGAGTTAAAGTGAAGACGAGGCAATGG 1435
QY 4039 TATTTATCGGCTGTGTCAGAAATCTACTGCTTTAAACCCAGCAGAGGGAACCCATAG 4098
Db 1436 TCTTTGGCAGCATTTGTTTCAGAGATCACTGTATTATCAAAAACAAGAAATAAATTTATTC 1495
QY 4099 AACTCTCAGGAGAGCCCTAGACCCACTTTTCATGGATCCAGACTTGGCATATGNACTT 4158
Db 1496 AAGATCCAGAAAAATAT---GATCCATTTATTCATGCACCCCTGATCTGTTCTGTGAACAC 1552
QY 4159 ATACAGGAAGCTGTGGTCTATGTAATGCACGAGTGTCTGGCAGAGAGTATTTTGAAGCTG 4218
Db 1553 ACATAGTAGCTGTGGCAGCATTTATGCATGCCCATTTGTTGGCAAGAGTATTTTGTATCG 1612
QY 4219 TACAGCTGAGCTCTCAGCAGCGCATTCATGTTGACCTTTT-----TGACTTGG 4266
Db 1613 TTCAAGCTAAAAGAACAGGCAAGGCAACAGAGATTACGCTTACATACGAGCTATGATGAG 1672
QY 4267 AAGTGGAGATATCTTTGCCCTCTTTCGAATCTCTGCAATCTCTGCAATCTGATGCCCATTA 4326
Db 1673 AAAACGGAGAAATCTTTTGGCCCCCTTTGTGAATGCTTGAGTAACTGTTTCTCTCTGCG 1732
QY 4327 TTCCTTTGCAACCTCAAAAGATTAACAGTGAAGATGCAGATGCTCTGCTGAACCTTTTGA 4386
Db 1733 TGCCTT-----CTCCAGAAATATTTTAAACACAGGTTAAATTTTTCAGACCAACCAA 1786
QY 4387 CCCCTGGCAGGTGGATACAGACTGTCTTGGCCAGAAATATCAGGTTATATAATAGACATG 4446
Db 1787 ATCTGACTCAGTGGATTAGAACAATATCTCAACAATAAAGCATTACAGTTTCTTAGGA 1846
QY 4447 CTAAGGAGAAACCCCAATTCCTATTTCTTTAATCAAGGAATGGAGATTTCTACTTGG 4506
Db 1847 AAGAAGAAGTACTCCTAATAATGCTCTACAAGAATTCAGAAATTTGGATGAATTAAC 1906
QY 4507 AGTTCCATTCCTGAGTGTGCGTTTGGCTTTCGATTAATTAATTAATTAATTAATTAATTA 4566
Db 1907 AGCTCCCTGAAGGTTTCCAGGCTGATTTTCGCTCTAAGATCCCTTATTTCTGAGAGCATAA 1966
QY 4567 AGGAAATGTTTATCTCTTTGGCCACAACAATTTATAGAAATTTGATTTGAAGTGGCCACTG 4626
Db 1967 AAGAAATGCTAACGACATTTGGAACTGCTACCTACAAGTGGGACTTAAAGTTTCATCCCA 2026
QY 4627 ATGAAAGGATTCCTCGAGTCCCATGCTGACCTGGAGACCTCGCGTTTCACTATCCAGG 4686
Db 2027 ATGAGAGAGCTCCTCGTGTCCCAATGTTGGGTAGCTGCGCGTACACCATCAAA 2086
QY 4687 CAATTGAAATCTATTGGAGATGAAGGAAAACCTCTGTTTGGAGCATTCAAAATAGGC 4746
Db 2087 GCATAGAAGAAATTTTGGATGATGAAGATAAACCATTTGTTGGTCTTTTACCTTGCAGAC 2146
QY 4747 AGCAATATGGTCTGAAAGCATTTAATGCAGTTTTCGATTTGCAGAGAGATTAACCTGCTC 4806
Db 2147 TGGATGACTGTCTTAGGTCAATTGACGAGATTTGGCCGACGACACTGGACAGTGGCATCAG 2206
QY 4807 AGGTCTCTATCAGAGAAACATCTGGTTCGCTCTATCAGTTGTTCTTCTCTAARATAAAT 4866
Db 2207 TTTTCAGTGGTGAAGGACATTTTGTGTAACCTTTTTCGACTGCTGCTGCTTAATGACAGCC 2266
QY 4867 CAGAAGATACACATGCTCTGCTATAGATCTGTTTTCATGTTTGGTGGGTGCTGTGT 4926
Db 2267 ATGAGGAACCTTCATGTCATATTAGATATTGACATGTTTTCATTTATTTGGTGGGTGCTG 2326

	QY	4927	TAGCATTTCCATCGCTTGTATTGGTAGACCCGTGTAATCNGCAGCCCTTCCTCAGTTAGTT	4986
	Db	2327	TTTTTTTTTCTCGGTGGTCACTG-----TCAGGATTTTTCAGGATCAGCC	2371
	QY	4987	CITCCATAACCACCTTTAATCTCTCCATTTGATCACCATGGCACACATGCTTCAGATAC	5046
	Db	2372	TTGGCACTGGAGACTTCACAATTTCCATCTGGTACTATGGCACACATCATACAGATCT	2431
	QY	5047	TACTTTACAGTAGACAC---AGGCCTACCCCCTTGCTCAGGTTCAAGAAGACAGTGAAGAGG	5103
	Db	2432	TACTTTACTCTCATGTACAGAAGAAATGGCATGGATCAAGAAAAATCCCCCTTGTGAAGAAG	2491
	QY	5104	CTCATTCGCCATCTCTCTTCTTGCGAGAAAATTTCTCAATATACAAGTGGCTCCATTGGGT	5163
	Db	2492	AATCAGAGTTCTTGTTGTATAAAACACTTCACCAGTATACGGGAAGTGCCTTGAAG	2551
	QY	5164	G TGATATTCCTGGCTGGTATTTCTGGGTCTFCAC TGAAGAAATGGCATCACCCC TTATCTTC	5223
	Db	2552	AAATACCATCCGCTGGCATCTCTGGANGAGTCTCAGAGCTGGAATCATGCCCTTCTGA	2611
	QY	5224	CTGTGCTGCAATGTTTTCCTCACTATTTACTTGGGTAACTCCGCCTGAGGAACATGCATA	5283
	Db	2612	AGTGTCTGCTTATTTTTCATTACTTAATAGAGTTCCTTCCCAACCCGACATCAAG	2671
	QY	5284	CCAATTCCTCAGAAGAGAGTACAGTGCACCTCTCTAGCTATCTATCTTTACCTTACAAAT	5343
	Db	2672	T----TCTTGGAACAGCCATTTTGAACTTTATGTAGCTATCTTTCCCTCAAACAACCC	2728
	QY	5344	TGTTCTCTGCTTTCCAGGAATATTTGGGATACTGTAAGGCCCTTGCTCCAGAGGTGGTGTG	5403
	Db	2729	TCATTTGCCCTTTTCAAGRAAATAGTGAGATAATGAATTCAC TGAATGAAAGTTGGTGCC	2788
	QY	5404	CAGATCGTGCCTTACTTAACATGTTTGAAGCAAAAACACGGTGGTCAGGTACCCFAGAA	5463
	Db	2789	GTAACAGTGAAGTTAAAGAGATATCTAGAAGGTGAAAGAGATGCTATAAGATATCCAAGAG	2848
	QY	5464	AAAGAAATAGTTGTATAGAGTCTCTGTAGTACATATAGTGCCTCCTGAATCAAGCTTCTC	5523
	Db	2849	AATCTAACAAATTAATAAACCTTCAGAGGATTTACAGCAGCCCTATTAATCAAGCATCCA	2908
	QY	5524	ATFTCAGGTGCCACGGTCTGCAGATGATGAGCGAAAGCATCTGTCTCTGTGCCTTTTCT	5583
	Db	2909	ATTTCTCGTCCCGAAATCAGGTGGTGATAAGAGCAGAGCCCCAACCTCTGTGCTTGTGT	2968
	QY	5584	GTGGGCTATACTATGTTCTCAGAAACATTTGTCGCCAGGAAATTTGARACGGGAAGAGG	5643
	Db	2969	GCGGATCTCTGCTGCTCCCAAGATTACTGTCTGCCAGACTGAACTTGAAGGGGGAGGATG	3028
	QY	5644	TTGGAGCTTGCAATTTTTCACGCATCTCCTGTGGACCGGAGTCTCATTTTCTCTAAAAA	5703
	Db	3029	TAGGAGCTGCAGACTCACACCTACTCTCTGTGGCTCTGGAGTGGGCATCTTCTCAGAG	3088
	QY	5704	TCAGAAATGCCAGTGGTCTGGTTGAAGGTAAAGCCAGAGGCTGTGCCTATCCAGCTC	5763
	Db	3089	TAGCGGAATGTCAGGTGCTATTTTtagtggcAAAAACCAAGGCTGTTTTTATCTCCTC	3148
	QY	5764	CTTACTTGNATGAATATGGAARACAGACCTGGCTTGAGAGGGGCGAACCCCTTCATTT	5823
	Db	3149	CTTACCTTGATGACTATGGGAGACCGACCGGGACTCAGACGGGGAATCCCTTTACAT	3208
	QY	5824	TATCTCGTAGGGCGTATCGGAAGCTCCATTTGTCCTGGCAACACAACTGATTATAGAAG	5883
	Db	3209	TATGCAAGAGCGATTAAGAAGATTCAGNAGCTCTGGCACCAACACAGTGTCCAGAGG	3268
	QY	5884	AGATTGCTAGGACCCAGAGACTTAATCAGATGTTATTTTGATTTCAACTGGCA	5935
	Db	3269	AAATTTGCAATGACAGGAGCAATCAGACACTGGTGGCATTTGACTTGCA	3320

RESULT 2

RESULT 2
US-10-414-378-56

US-10-414-378-38
; Sequence 56, Application US/10414378

; Sequence ID, Application US/104
; Publication No. US20030165981A1

; GENERAL INFORMATION:									
; APPLICANT: FUKUSHIMA, DAIKICHI									
; APPLICANT: SHIBAYAMA, SHIRO									
; APPLICANT: TADA, HIDEAKI									
; TITLE OF INVENTION: POLYPEPTIDE, CDNA ENCODING THE POLYPEPTIDE, AND USE OF									
; TITLE OF INVENTION: THE BOTH									
; FILE REFERENCE: Q38769									
; CURRENT APPLICATION NUMBER: US/10/414,378									
; CURRENT FILING DATE: 2003-04-16									
; PRIOR APPLICATION NUMBER: US/09/529,063									
; PRIOR FILING DATE: 2000-04-07									
; PRIOR APPLICATION NUMBER: PCT/JP98/04514									
; PRIOR FILING DATE: 1998-10-06									
; PRIOR APPLICATION NUMBER: JP 9-274674									
; PRIOR FILING DATE: 1997-10-07									
; NUMBER OF SEQ ID NOS: 117									
; SOFTWARE: PatentIn Ver. 2.1									
; SEQ ID NO 56									
; LENGTH: 3327									
; TYPE: DNA									
; ORGANISM: Homo sapiens									
US-10-414-378-56									
Query Match 12.9%; Score 813.6; DB 12; Length 3327;									
Best Local Similarity 55.4%; Pred. No. 7.9e-215;									
Matches 1856; Conservative 0; Mismatches 1409; Indels 87; Gaps 11;									
Qy	2632	TGGAATATCCTTTACGTTGTCTGGTGTGCTTGTGCTGCCAGGTTGTGCTGAGATGTGGCGAA	2691						
Db	8	TAGAACACCCTCTAGATGCTTGTCTGTGTGCCAAGTACATGCCGGAATGTGGAGAA	67						
Qy	2692	GAATGACATGCTCTCTATTAGCCAGCGTGTATTACCAAGATGTTAACTGCAGAGAAG	2751						
Db	68	GAATGGGTTCTCTCTAGTAGTAACAGATTTATTACTACCATAATGTCAATGCAGACGCTG	127						
Qy	2752	AAATGATGATAAAGATATCATCGTCTCAGATTTGGTGCATCTTTAATGGATCCCAATA	2811						
Db	128	AGATGTTTGCAAGGATGTAGTAATGCTTCAGACAGGTGTCTCCATGATGGATCCAAATC	187						
Qy	2812	AGTCTGTGTACTGGTACTTCAGAGGTPATCAACTGCCCGAGGCTTTTAAACAAGCCATAT	2871						
Db	188	ATTCTCTGATGATCATGCTCAGCGCTTTGAACTTTATCAGATTTTACGTACTCCAGACT	247						
Qy	2872	CTACAAA-----GACCAGGATTTGATTAACAATATAATACAC	2910						
Db	248	ATGAAAAGATTTAGTCTTGATATACCATAAGGATGTGTTACGAGAACATACTC	307						
Qy	2911	TAATAGAAGAAATGCTTCAGGTCCATCATATATTGGGTGAGCGGTTATGTACCTGGAG	2970						
Db	308	TAATAGAAGAAATGCTATACCTCATTTAATGCTTGTGGAGAGAGATTTAGTCTCGGAG	367						
Qy	2971	TGGGAATGTGCAAGAAGAGGTACATGAGAGAAATCATCTACTGCTTTGCATTG	3030						
Db	368	TTGGCAGGTAAATGCTACAGATGAAATCAAGCGAGAGATTTATCCATCAGTTAGATATCA	427						
Qy	3031	AACTGTCACACAGTGCCATTGCCAAAAATTTACTGAGATGAAATTAATGAACATG	3090						
Db	428	AGCTATGGCTCATAGTGAATTCGTAAGTCTTTACTGAGATGAGACACAGGACACTG	487						
Qy	3091	GCTTAGAGAATGTCATAAAACAAAGTGCCACATTTAAGAAACACAGGTGTATCAGGCCATG	3150						
Db	488	GCATGGAGAGTGTAATCGAAGACGATGGCCCATTTCAAGAAACCTGGATTAAACAGGACGAG	547						
Qy	3151	GAGTTTATGAACATAAAGATGAATCACTGAAAGACATTCATATGTAATCTTTATCATTA	3210						
Db	548	GCATGATGAACATGAAACCAAGATGTGCCAAGAGTTCAACTTGTATTTCTATCACTTTT	607						
Qy	3211	CCAAACCCAGCATAGCAGGCTGAACATATGAGAGAAAGAGCAAAACAAAGAAACA	3270						
Db	608	CAAGGGCAGAACAGTCCAGGGCAGAGAGACGCGCAACGGAATTTGAAAGACAAATATAG	667						
Qy	3271	AAGATGAAGCATTTGCCGCCACCACCACTCTCTGAATTTCTGCCCTGCTTTTCAGCAAAAGTGA	3330						

Db 568 AAGATACAGACTCCACCTCCGGTGTGGCTCCATTCTGCCCTCTGTGGCAAGCCTGG 727
QY 3331 TTAACCTTCTCAACTGTGATATCATGATGATACATTTCTCAGGACCGTATTTGAGCGGCA 3390
Db 728 TTAACATTTTGGAGTCAGATGTCATGTGTGTCATGATGGAACAATTTCTGCAATGGGCTG 787
QY 3391 TAGACAGATTTCACTTGTGGACCGGAAGGATGCTCCAAATGGCTTTTCATATTTCTGG 3450
Db 788 TGAACATATGATGATGCTGTCAGAGTCCATGCTGCAAGAGGGTGTACATTTAATG 847
QY 3451 CATTGGTTTACTAGAGAGAACCAACAGCTTCAAAAAGCTCTCTGAAG---RAGAAGTAA 3507
Db 848 GCATGGCATACAGAGAAACAAACATTTAGAGATGTCACGGAAGCATGTAGTAA 907
QY 3508 CATTGGACTTTTATCATAGGCTTCAAGTTGGGAAGTTTCAGGCATGAAT-----A 3558
Db 908 CATTTACCTTCACTCAGAGATATCAAAACCTGGTGAAGGCCCAAAAATTTCTCCTAGCA 967
QY 3559 TACAATGCTTTTGGAAAACTCAAGGAATTTCCCAAGTTAGAGGCCGAAGACATGA 3618
Db 968 TACTAGCTATGCTGGAACACTACAAAATGCTCCCTACCTAGAGTCCCAAGACATGA 1027
QY 3619 TAACGTGATCTCAGATGTTTGACACAGTGAAGCATTAAGAGAAAATCTGTGTTAA 3678
Db 1028 TTCGGTGATATGAAGACTTTTAATGCTGTTAAAAAGATGAGGAGA-----GTTTCA 1081
QY 3679 TTGTAGCAACACATCAGGATCGGAATCTATTAGAAATGATGAGATTAATCATGATAAG 3738
Db 1082 CTACCACTCCGTTGGCAGACAGAGGAAACCATTAATGGAAGAGTTCAAGGGACAAG 1141
QY 3739 AAAAGCAGAACGAAAAAGCTGAAGCTGCTAGGCTACATTCGCCGAAGATCATGG 3798
Db 1142 ACAAGCTGAGAGGAGAAAAAGCAGAGATGCGACACTGCGCAGAGAAAAAGATCATGG 1201
QY 3799 CTCAGATGCTGCTTACAGAAAAAATTCATTGAACTCATAACTCATGTATGACATA 3858
Db 1202 CTCAGATGCTGAAATCGCGGCAATTTATTGATGAAAAACAAGAACTCTTTTCAGCAGA 1261
QY 3859 CATCAGAAATGCTGGGAAAGAGATTCATATGAGGAGAGAGACAGCCACGACATCA 3918
Db 1262 CATTAGAACTGGATGGCTCAACCTCTGCTGTTTGTATCA-----TAGCCCTGGGCTT 1315
QY 3919 GTGACTACTCTAGAAATTTGCTTTGGGTCTTAAACGGGTCTCATCTGTTACTGAAAGGAG 3978
Db 1316 CAGATATGACACTTACAGCACTGGGCCCGCACAACTCAGGTTCTTGAACAAGACAAT 1375
QY 3979 TGTGACGTGATCCTTTGGCAGAGAACAGAGAGGTGAATAAGAAAAATATCCATGG 4038
Db 1376 TCGTTACATGATATTTGTGTCAGAGGAGCAAGAAGTTAAAGTGAAGCAGGGCAATGG 1435
QY 4039 TATTATCGGCTGTGCCAGAAATCTACTGCTTTAACCCAGCAGAGGGGAAACCCATAG 4098
Db 1436 TCTTGGCAGCATTTGTTTCAGAGATCAACTGATGATTTATCAAAAACAGAGTAAATTTATC 1495
QY 4099 AACTCTCAGGAGAACCCCTAGACCCACTTTTCATGGATCCAGACTTGGCATATGGAACCT 4158
Db 1496 AAGATCCAGAAAATAT---GATCCATTATCATGCAACCTGATCTGTTGTGGAACAC 1552
QY 4159 ATACAGGAAGCTGTGGTCATGTAATGACGAGGTGCTGCGCAGAGATTTTGAAGCTG 4218
Db 1553 ACACTAGTAGCTGTGGGCATATTATGATGCCATTTGTTGCAAGGATTTTGTATTCGG 1612
QY 4219 TACAGCTGAGCTCTCAGCAGGCAATTCATGTTGACCTTTT-----TGACITGG 4266
Db 1613 TTCAAGCTTAAGAAACAGCAGGAAACACAGAGATTAACGCTTACATACGAGCTATGATGAG 1672
QY 4267 AAAGTGAGAAATATCTTTGGCCTCTTTGCAAAATCTCTGTGCAATACTGTGATCCCATTA 4326
Db 1673 AAAACGAGAAATCTCTTTGGCCCCCTTTGTGAATGCTTGAGTAATACTGTTATTCCTGCG 1732
QY 4327 TTCCTTTGCAACCTCAAAAGATAAACAGTGAAGATGCTGCTCTGCTCAACTTTTGA 4386
Db 1733 TGCCTT-----CTCCAAGAAATATTTTAAACAACAGGTTTAAATTTTTCAGACCAACCAA 1786

QY 4387 CCCTGGCACGGTGGATACAGACTGTTCTGGCCAGAAATATCAGGTTATAATATAGACATG 4446
Db 1787 ATCTGACTGATGATTAGAACAAATATCTCAGCAAAATAAAGCATTTACAGTTTCTTAGGA 1846
QY 4447 CTAAGAGAAAAACCCAAATTCCTATTCTTTTAATCAAGGAATGGAGATTTCTACTTTGG 4506
Db 1847 AAGAAGAAAGTACTCCCTAATAATGCCCTCTACAGAAATTCAGAAATTTGGATGAATATAC 1906
QY 4507 AGTTCCATTCCTAGTTTGGCGTTTGGCTTTCGATTTAAATATTTCAATAGCATCA 4566
Db 1907 AGTCCCTGAAGGTTTCAAGGCTGATTTTCGCTCTAAGATCCCTTATTTCTGAGACATAA 1966
QY 4567 AGGAAATGGTTATTTCTTTGGCCACAACAATTTATAGAAATTTGGATTTGAAAGTGCACCTG 4626
Db 1967 AAGAAATGCTAACGACATTTTGAACCTGCTACCTACAAGTGGGACTAAAGTTTATCCCA 2026
QY 4627 ATGAAAGGATCTCGAGTCCCATGCTGACCTGGAGCAGCTGGCTTTTCACTATCCAGG 4686
Db 2027 ATGAAAGGATCTCGTGTTCCTCATATGTTGGGGTAGCTGCGCGTACACCATCCCAA 2086
QY 4687 CAATTGAAAATCTATTGGGAGATGAAGAAAAACCTCTCTTTGGAGCACTTCAAAATAGGC 4746
Db 2087 GCATAGAAAGATTTTGGTGTGATGAAGATAAACCATTTGTTGGTCTTTTACCTTGCAGAC 2146
QY 4747 AGCAATATGGTCTGAAAGCATTAATGCAAGTTTGGAGTTGACAGAGATTTACCTGTCCTC 4806
Db 2147 TGGATGACTGCTTAGGTCAATTCAGAGATTTTGGCGCAGCACACTGGACAGTGGCATCAG 2206
QY 4807 AGGTCCTGATACAGAAACATCTGGTCTCTCTATCATCAGTTGTTCTTCTTCACTAAATAAT 4866
Db 2207 TTTCAAGTGGTCAAGGACATTTTGTAAACCTTTTTCATCACTGCTGCTCAATGACAGCC 2266
QY 4867 CAGAAGATACACCATGCTCTCTCTATAGATCTGTTTTCATGTTTGGTGGGTGCTGTGT 4926
Db 2267 ATGAGAACTTCCATGCAATATAGATTTGACATGTTTCAATTTATTTGGTGGCTTGGTGC 2326
QY 4927 TAGCATTTCCCATCTCTTATTTGGGATGACCCCTGTTGATCTGCAAGCTTTCTTCACTAGTT 4986
Db 2327 TTGCATTTCTCGGTGTCAGTG-----TCAGGATTTTTCAGGATCAGCC 2371
QY 4987 CTTCCTATAACACCTTTATCTCTTCCATTTGATCACCATGGCACACATGCTTCAGATAC 5046
Db 2372 TTGGCACTGGAGACCTTCCATTTTCCATCTGGTTACTATGGCACACATCATACAGATCT 2431
QY 5047 TACTTACAGTAGACAC-----AGCCCTTACCCCTTGTCTCAGTTTCAAGAAGACAGTGAAGAG 5103
Db 2432 TACTTACCTCATGTACAGAGAGAAATGGCATGGATCAAGAAATCCCCCTTGTGGAAGAG 2491
QY 5104 CTCATTTCCGCATCTTTCTTTTTCAGAAATTTCTCAATATACAAGTGGCTCCATTTGGGT 5163
Db 2492 AATCAGCAGTTCTTCTGTTTGTATAAACAACATTCACCAAGTATACGGGAAGTGCCTTGAAG 2551
QY 5164 GTGATATTTCCCGGTGGTATTTTGGGTCTCTACTGAAGAAATGGCATCAGCCCTTATCTTC 5223
Db 2552 AAATACCATCCGGCTGGCATCTGTGGAGAGTGTGAGAGTGGAAATCATGCTTCTCCCTGA 2611
QY 5224 GCTGTGCTGATTTGTTTTCACATTTTACTTGGGTAACTCCGCTGAGGAACATGCATA 5283
Db 2612 AGTGTCTGCTTATTTTTCATTTAAATGGAGTTTCCCTTCCCAACCCAGACATTCAG 2671
QY 5284 CCAATTTCTGAGAGAGAGATACAGTGCATCTGTAGCTATCTATCTTTTACCTACAAATTT 5343
Db 2672 T---TCCCTGGACAGACCATTTTGAACATTTATGATGCTATCTTCCCTTACCAACAACC 2728
QY 5344 TGTTCCTCTCTTCCAGGAATATTTGGGATCTGTAAGGCCCTTGTGCTCCAGAGGTGGTGTG 5403
Db 2729 TCATTTGCTTTTTCAGAAAAATAGTGAGATAATGAATTTCACTGATTTGAAAGTTGGTGC 2788
QY 5404 CAGATCCTGCTTACTAACTGTTTGAAGCAAAAAACACCGTGGTTCAGCTACCCCTAGAA 5463
Db 2789 GTAACAGTGAAGTTAAAAAGATATCTAGAGGTGAAGAGATGCTATAAGATATCCAAAG 2848

QY 5464 AAAGAAATAGTTTGTATAGAGCTTCTGTATGACTATATAGCTGCCTCTGAATCAAGCTTCTC 5523
DB 2849 AATCAACAAATTAATAAATCTCCAGAGGATTACAGCAAGCTCAATTAACAGCATCCA 2908
QY 5524 ATTTCAGGTGCCAGGCTGTCAGATGATGAGCGAAAGCATCTGTCCTCTGCTTTTCT 5583
DB 2909 ATTTCCTGTCGCCGAAATCAGGTGTTGATAAGAGCAGAGCCCAACTCTGTGCTTGTGT 2968
QY 5584 GTGGGGCTATATCTTCTCAGAACATTTGCTGCCAGGAATTTGTAACGGGGAAGAG 5643
DB 2969 GCGGATCTCTGCTGTGCTGCCAGAGTTACTGTGCCAGACTGAACCTGGAAGGGGAGGATG 3028
QY 5644 TTGGAGCTTGCAATTTTTCAGGCACCTCACTGTGGAGCCGAGTCTGCATTTTCTCTAAAA 5703
DB 3029 TAGGAGCTGACAGCTCACACTACTCTGTGGCTCTGGATGGGCATCTTCTCTGAGAG 3088
QY 5704 TCAGAGAAATGCGAGTGTGCTGTTGAAGGTAAAGCCAGAGCTGTGCTATFCCAGCTC 5763
DB 3089 TACGGGAATGTGAGGTGCTATTTTGTAGCTGGCAAAACCAAGGCTGTTTTATTCTCCTC 3148
QY 5764 CTTACTTGGATGAATATGAGAGAAACAGACCTTGGCTGGAAGGGGCAACCCCTTCATT 5823
DB 3149 CTTACTTGTGACTATGCGGAGACCGACACGGGACTCAGACGGGGAATCTCTTTACATT 3208
QY 5824 TATCTGCTGAGCGGTATCGGAAGCTCCATTTTGTGCTGGCAACAACTGCATTATAGAAG 5883
DB 3209 TATGCAAGAGCGGATTCAGAAAGATTTCAGAACTCTGGCACCACACAGTGTACAGAGG 3268
QY 5884 AGATTGCTAGGAGCCCAAGAGACTAATCAGATGTTATTTGGATTCAACTGGCA 5935
DB 3269 AAATTTGGACATGCACAGGAGCCCAATCAGACACTGTTGGCATTGACTGGCA 3320

RESULT 3

US-09-529-063-57
; Sequence 57, Application US/09529063
; Patent No. US20020102542A1
; GENERAL INFORMATION:
; APPLICANT: FUKUSHIMA, DAIKICHI
; APPLICANT: SHIBAYAMA, SHIRO
; APPLICANT: TADA, HIDEAKI
; TITLE OF INVENTION: POLYPEPTIDE, CDNA ENCODING THE POLYPEPTIDE, AND USE OF
; TITLE OF INVENTION: THE BOTH
; FILE REFERENCE: Q58769
; CURRENT APPLICATION NUMBER: US/09/529,063
; CURRENT FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: PCT/JP98/04514
; PRIOR FILING DATE: 1998-10-06
; PRIOR APPLICATION NUMBER: JP 9-274674
; PRIOR FILING DATE: 1997-10-07
; NUMBER OF SEQ ID NOS: 117
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 57
; LENGTH: 3502
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: (56)..(109)
; NAME/KEY: mat_peptide
; LOCATION: (110)..(3382)
; NAME/KEY: CDS
; LOCATION: (56)..(3382)
US-09-529-063-57

Query Match 12.9%; Score 813.6; DB 10; Length 3502;
Best Local Similarity 55.4%; Pred. No. 8.2e-215;
Matches 1856; Conservative 0; Mismatches 1409; Indels 87; Gaps 11;
QY 2632 TGGAAATATCTTTACGTGCTGTTGGTTGCCAGGTTGTGCTGAGATGTGCGGAA 2691
DB 63 TAGAACACCCCTCTTAGATGCTCTGTTCTGTGTGCCCAAGTACATGCCGGAATGTGGAGAA 122

QY 2692 GAAATGACTGCTCTCTTATTAGCCAGGTGTTTATTACCAAGATGTTTAAAGTCACAGAG 2751
DB 123 GAAATGGGTCTCTCTTAGTAAACAGATTTTATTACTACCAATAATGTAAGTGCAGACGTG 182
QY 2752 AAATGTATGATAAAGATATCATCTGCTTCAAGATTGGTGCATCTTAAATGATGCCAATA 2811
DB 183 AGATGTTTGACAAGGATGTAGTAATGCTTTCAGACAGGTGCTCTCCATGATGGATCCAAATC 242
QY 2812 AGTTCTTGTACTGGTACTTCAGAGGTATGAACCTTGGCCAGGCTTTTACAAAGACCATAT 2871
DB 243 ATTCTCTGATGATCATGCTCAGCCGCTTTGAACCTTTATCAGATTTTCTAGTACTCCAGACT 302
QY 2872 CTACAAAA-----GACCAGGATTGATTAAACAAATATATACAC 2910
DB 303 ATGGAATAAGATTTAGTTCGAGATTACCCATAGGATGTTGTCAGCAGAACAAATCTC 362
QY 2911 TAATAGAAGAAATGCTTCAGGTCTCTCATCTATATTGTTGGTGAGCGTTTATGTACCTGGAG 2970
DB 363 TAATAGAAGAAATGCTATACCTCATTTAATGCTTGTGAGAGAGATTTAGTCTCTGGAG 422
QY 2971 TGGGAATGTGACCAAGAGAGGTCAATATGAGAGAAATCATTCACATTGCTTTGCTATTG 3030
DB 423 TTGCACAGGTAAATGCTACAGATGAAATCAAGCGAGAGATTTATCCATCAGTTGAGTATCA 482
QY 3031 AACCCATGCCACACAGTGCCTTCCCAAAATTTACCTGAGAAATCAAAATATGAACATG 3090
DB 483 AGCCTATGGCTCATAGTGAATTTGTAAGTCTTTACCTGAAATGAGAACAGAGAGACTG 542
QY 3091 GCTTAGAAGATGTCTAAACAAAGTGGCCACATTTAAGAAACAGAGTGTATCAGGCCATG 3150
DB 543 GCATGGAGAGTGTAAATCGAAGCAGTTGCCCATTTTCAAGAACTGGATTAACAGCAGCGAG 602
QY 3151 GAGTTTATGAACATAAAGATGAATCACTGAAAGACTTCAATATATGACTTTTATCATTTACT 3210
DB 603 GCATGTATGAATGAAACCAAGATGTGCCAAGAGTTCAACTGTTGTTTCTATCATCTTT 662
QY 3211 CCAAAACCCAGCATAGCAAGCTGAACATATGCAGAAAGAGAGAGAAACAGAAACCA 3270
DB 663 CAAGGGCAGAACAGTCCCAAGGCGAAGAACGCGAAATTTGAAAGACAAAATAGAG 722
QY 3271 AAGATGAAGCATTCGCCACACACCTCCTGTAATCTGCTCTGCTCTTTCAGCAAAAGTGA 3330
DB 723 AAGATACAGCACTCCCACTCCGCTGTTGCTCTCATCTGCTCTGCTCTGTTTGAAGCCTGG 782
QY 3331 TTAACCTTCTCAACTGTGATATCATGATGATCATCTTCTCAGGACCGTATTTTGAAGCGGCA 3390
DB 783 TTAACATTTTGCAGTCAGATGCTATGTTGTCATCATGGAACAATTTCTGCAATGGCTG 842
QY 3391 TAGACACAGATTTCAACTTTGGACCGAGGATGCTCCAAATGCTTTTCAATATCTG 3450
DB 843 TGGAAACATAATGGATATGCTCTGTCAGAGTCCATGCTCAAAAGGCTGTACATTTAATTG 902
QY 3451 CATTGGGTTTACTAGAAGAGAGCAACAGCTTCAAAAAGCTCTCTCAAG---AAGAAGTAA 3507
DB 903 GCATGGCACTACAAGAGAGAAACAAACATTTAGAGAATGTCACGGAAGAGCATGATGATA 962
QY 3508 CATTGTGCTTTTATCATAAAGCTTCAAGATTGGGAAGTTCAGCCCATGAAT-----A 3558
DB 963 CATTTCCTTCATCAGAAAGATATCAAACTGTTGAAGCGCCCAAAAATTTCTCTTAGCA 1022
QY 3559 TACAAATGCTTTTGGAAAACTCAAGGAATTTCCCAAGTTTGAAGCGCCAGAGGACATGA 3618
DB 1023 TACTAGCTATGCTGGAACACTACAAAATGCTCCCTACCTAGAGTCCACAAAGACATGA 1082
QY 3619 TAACGTGGATACTTCAGATGTTTGACACAGTGAAGCGATTAAAGAGAAAAATCTTGTGTTAA 3678
DB 1083 TTCGGTGGATATTGAAGACTTTTAAATGCTGTTAAAAAGATGAGGAGA-----GTTTAC 1136
QY 3679 TTGTAGCAACCATCAGGATCGGAATCTATTAGAATGATGAGATTTACTCATGATAAG 3738
DB 1137 CTACCACTCCCGTGGCAGAGACAGAGGAACCAATTAATGGAAGAGAGTTTCAAGGGCAAG 1196
QY 3739 AAAAGCAGAACGAAAAAGAAAGCTGAAGCTGCTAGGCTACATGCCAGAGAGATCATGG 3798

RESULT 4

```

US-10-414-378-57
; Sequence 57, Application US/10414378
; Publication No. US20030165981A1
; GENERAL INFORMATION:
; APPLICANT: FUKUSHIMA, DAIKICHI
; APPLICANT: SHIBAYAMA, SHIRO
; APPLICANT: TADA, HIDESAKI
; TITLE OF INVENTION: POLYPEPTIDE, CDNA ENC
; TITLE OF INVENTION: THE BOTH
; FILE REFERENCE: Q58769
; CURRENT APPLICATION NUMBER: US/10/414,378
; CURRENT FILING DATE: 2003-04-16
; PRIOR APPLICATION NUMBER: US/09/529,063
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: PCT/JF98/04514
; PRIOR FILING DATE: 1998-10-06
; PRIOR APPLICATION NUMBER: JP 9-274674
; PRIOR FILING DATE: 1997-10-07
; NUMBER OF SEQ ID NOS: 117
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 57
; LENGTH: 3502

```

Query Match	12.9%;	Score 813.6;	DB 12;	Length 3502;
Best Local Similarity	55.4%;	Pred. No. 8.2e-215;		
Matches 1856;	Conservative	0;	Mismatches 1409;	Indels 87; Gaps 11;
QY	2632	TGGAATATCCGTTTACGTTGTCTCGGTGTGGTTGCCCAGGTTGTTGCTGTGAGATGTGGCGAA	2691	
DB	63	TAGAACACCCTCTTAGATGTCTTGTGTGTGCCCAAGTACATGCCGAATGTGGAGAA	122	
QY	2692	GAATGSGACTGCTCTCTTATTAGCCAGGTGTTTATTACCAAGATGTTTAAGTCGAGAGAAG	2751	
DB	123	GAATGGGTCTCTCTTAGTAACACGAGTTATTATTACCAATATGAAATGAGACGCTG	182	
QY	2752	AAATGTATGATAAAGATATCATCATGCTTCAGATTGGTGCATCTTTAATGGATCCCAATA	2811	
DB	183	AGATGTTTGACAGGATGTAGTAATGCTTCAGACAGGTGTCCTCATGTGGATCCAAATC	242	
QY	2812	AGTCTGTTTACGGTACTTCAGAGGTATGAACCTGCCGAGGCTTTTAAACAAGACCATAT	2871	
DB	243	ATTTCCTGATCATGCTCAGCGCGTTTGAACCTTTATCAGATTTTTCAGTACTCCAGACT	302	
QY	2872	CTACAAAA-----GACCAGGATTTGATTAACAATATAATACAC	2910	
DB	303	ATGGAATAAGATTAGTCTCTGAGATTACCCATGAAGATGTTGTTTCAGCAAGAACATACTC	362	
QY	2911	TAATACAAGAAATGCTTCAGGTCCCTCATCTATATTGTTGGGTGAGCGGTTATGTACCTGGAG	2970	
DB	363	TAATAGAAGAAATGCTATACCTCATTTAATGCTTGTGGAGAGAGATTTAGTCTCGGAG	422	
QY	2971	TGGGAATGTGACCAAGAAGAGGTCACAAATGAGAGAAATCATTTCACTTGCTTTGCATTG	3030	
DB	423	TTGGACAGGTAAATGCTACAGATGAATCAAGCGAGAGATTATCCATCAGTTCAGTATCA	482	
QY	3031	AACCCATGCCACAGTGGCAATTGCCAAAAATTTTACCTGAGAAATGAAAATTAATGAACCTG	3090	
DB	483	AGCCTATGCTCATACTGAATTTGGTAAGTCTTTTACTCTGAAGATGAGACAAGAGACTG	542	

QY	3091	GCTTAGAAGATGTCATAAAACAAAGTGGCCACATTTTAGAAAACCGAGTGTATCAGGCCATG	3150
Db	543	GCATGGAGAGTGTAAATCGAAGCAGTTGCCCATTTCAAGAAACCTGGATTAACAGGACGAG	602
QY	3151	GAGTTTATGAACATAAAGATGAATCACTGAAAGACTTCAATATGTACATTTTATCATTA	3210
Db	603	GCATGTATGAACGTGAACACCAAGATGTCCAAAGAGTTCAACTGTGATTCTTATCACTTTT	662
QY	3211	CCAAAACCCAGCATAGCAAGGCTGAACATATGCAGAGAAAGGAGAAAACAAGAAAACA	3270
Db	663	CAAGGCCACACAGTCCAAGGCAGAGAGCGCAACGGAATTTGAAAGACAAATATAGAG	722
QY	3271	AAGATGAAGCATTCGCCGCCACCAACCACTCCTGAATTCCTGCCCTCTTTCAGCAAAAGTGA	3330
Db	723	AAGATACAGCACTCCCACTCCGGTGTGCTCCATCTGCTCCATCTGCTGTTTGAAGCGCTGG	782
QY	3331	TTAAACCTTCTCAACTGTGATATCATGATGTACATTTCTCAGAGCCGTATTTGAGCGGGCAA	3390
Db	783	TTAAACATTTGCACTCAGATGTGATGTGTGCATCATGGGAACAATTTCTGCAATGGCGTG	842
QY	3391	TAGACACAGATTCTAACTTGTGACCCGAAGGATGCTCCAATGSCCTTTTTCATATCTCG	3450
Db	843	TGGAACATTAATGGATATGCTGTGCAGAGTCCATGCTGCAAGGGTGTACATTTAATTG	902
QY	3451	CATTGGGTTTACTAGAAGAGAACCAACAGCTTCAAAAGCTCTCGAAG---AAGAGTAA	3507
Db	903	GCATGSCACTACAAGAGAAAACAACATTTAGAGAATGTCACGGAAGCATGTAGTAA	962
QY	3508	CATTTGACATTTATCATAGGCTTCAAGATTTGGAAAGTTTCAGCCATGAAT-----A	3558
Db	963	CATTTTACCTTTCACTCAGAAGATATCAAAACCTGGTGAAGCGCCAAAAAATTCCTC	1022
QY	3559	TACAAATGCTTTTGGAAAACTCAAAGGAATCCCAAGTTAGAGCCAGAGGACATCA	3618
Db	1023	TACTAGCTATGCTGGAAACACTACAAATGCTCCCTACCTAGAAGTCCACAAAGACATGA	1082
QY	3619	TAACTGGGATCTTCAGATGTTTGACAGTGAAGCGATTAAAGAGAAAAATCTTGTTTAA	3678
Db	1083	TTCCGTGGATATTGAAGACTTTTAATGCTGTTAAAGATGAGGGAGA-----GTTAC	1136
QY	3679	TTGTAGCAACACATCAGGATCGGAATCTATTAGAATGATGAGATTACTCATGATAAAG	3738
Db	1137	CTACCAGTCCCGTGGCAGACAGAGAGGAACCAATAATGGAAGAGAGTTTCAAGGGACAAAG	1196
QY	3739	AAAAAGCAGAAACGAAAAAGCTGAAGCTGCTAGGCTACATCCCGCAGAGATCATGG	3798
Db	1197	ACAAAGCTCAGAGGAAGAGAAAGACAGAGATGCGCAGACTCGCAGAGAAAGATCATGG	1256
QY	3799	CTCAGATGCTGCCTTACAGAAAAACTTCATTGAAACTCATAACTCATGATGACAATA	3858
Db	1257	CTCAGATGCTGNAATTGCAGCGCATTTTATTGATGAAACAAAGAACTCTTTCAGCAGA	1316
QY	3859	CATCAAAATGCTTGGGAAGAAGATTCCATTATTGAGGAAGAGAGCACCACGACGTCA	3918
Db	1317	CATTAGAACTGGATGCTCAAACCTGCTGCTCTTGATCA-----TAGCCCTGTGGCTT	1370
QY	3919	GTGACTACTCTAGAATTGCTTTGGTGCTTAAACGGGGTCCATCTGTTACTGAAAAGGAGG	3978
Db	1371	CAGATATGACACTTACAGCACTGGGGCCCGCACAAACTCAGGTTTCTTGAACAAAGACAAT	1430
QY	3979	TGCTCAGCGCATCTTTGCCAAGAAGAACGAGGTTGAAATAGAAATAATGCCATGG	4038
Db	1431	TGTTTACATGTATTTGTGTCAGAGGAGCAGAAGTTTAAGTGGAAAGCAGGCAATGG	1490
QY	4039	TATTATCGCCCTGTGTCCGAAATCTACTGCCTTAAACCCAGCACAGCGGGAACCCATAG	4098
Db	1491	TCTTGGCAGCAATTTGTTTCAGAGATCACTGTATTATCAAAAAACAGAAGTAAATTTATTC	1550
QY	4099	AACTCTCAGGAGAGCCCTTAGACCCACTTTTTCATGGATCCAGACTTGTCATATGGAAT	4158
Db	1551	AAGATCCCAAAAAATAT---GATCCCATTTATCATGCACCCGATGCTGCTCTTTGGAAAC	1607

QY 4159 ATACAGGAAGCTGTGTCATGTAATGACGAGTGTCTGCCAGAAAGTATTTTGAAGCTG 4218
Db 1608 ACCTAGTAGCTGTGGGACATTTATTTGATGCTCCCATTTGTTGCAAGAGTATTTTGAATCCG 1667
QY 4219 TACAGCTGAGCTCTCAGCAGCGGATTCATGTTGACCTTTT-----TCAGTTGG 4266
Db 1668 TTCAAGCTAAGAACAGGAGGACACAGAGATTACGCTTACATACAGAGCTATGATGTAG 1727
QY 4267 AAGTGGAGAAATATCTTTGGCCCTTTTGGCAAAATCTGTGCAATPACTGTGATGCCCAATTA 4326
Db 1728 AAAACGGAGAAATCTCTTTGGCCCTTTTGTGAATGCTTGTGAGTAATGATGTTATTCCTCTGC 1787
QY 4327 TTCTTTGCAACCTCAAAAGATAAACAGTGAAGATGAGATGCTGCTTCTGCTCAACTTTTGA 4386
Db 1788 TGCTTT-----TCCCAAGAAATATTTTAAACACAGGTTAAATTTTTCAGACCAACCAA 1841
QY 4387 CCCTGGCAGCGGTGATACAGACTGTTCTGGCCAGAATATCAGGTTAATAATATAAGACATG 4446
Db 1842 ATCTGACTAGTGGATTAGAACATATCTCAGCAAAATAAAGCAATTACAGTTTCTTAGGA 1901
QY 4447 CTAAGGAGAAAACCCAAATTCCTATTTCTTTTAAATCAAGGAATGGGAGATTCTACTTTGG 4506
Db 1902 AAGAAGAAAGTACTCTCTAATAATGCCCTACAAAGAAATTCAGAAATGTTGGATGAATTAC 1961
QY 4507 AGTTCCATTCCCTCAGTTTGGCGTTGAGTCTTCGATTAATATTAATGATGATCA 4566
Db 1962 AGCTCCCTGAAGGGTTCCAGGCGTGAATTTCTGCTAAGATCCCTTATTTGAGAGCATAA 2021
QY 4567 AGGAATAGTTATTTCTTTGGCCACAACAATTTATAGAATTTGGAATGAAAGTGCACCTG 4626
Db 2022 AAGAAATGCTAACGACATTTGGAACCTGCTACCTACAGGTGGGACTAAAGTTTCTATCCCA 2081
QY 4627 ATGAAGGGATCCTCAGTCCCATGCTGACCTGAGGACACCTGCTGCTTTTCTACTATCCAGG 4686
Db 2082 ATGAAGAGGATCCTCGTGTCCCATATGTTGGGGTAGCTGCGCGTACACCATCCAAA 2141
QY 4687 CAATTGAAATCTATTGGGAGATGAAGAAACCTCTGTTGGAGCAGCTTCAAAATAGGC 4746
Db 2142 GCATAGAAAGAAATTTGAGTGATGAAGATAAACCATTGTTGGTCTTTTACCTTTGCAGAC 2201
QY 4747 AGCATAATGTTCTGAAAGCATTAAATGAGTTTGCAGTTGACAGAGGATTACCTGTCTC 4806
Db 2202 TGGATGACTGCTTAGTCTATGACGAGATTGCGCGAGCACACTGACAGTGGCATCAG 2261
QY 4807 AGTCTCTGATACAGAAACATCTGCTGCTCTCTATCAGTTGTTCTTCTTCAACATAAAAT 4866
Db 2262 TTTCAAGTGGTCAAGGACATTTTGTAACTTTTTCATCCTGCTGCTTAATGACAGCC 2321
QY 4867 CAGAAATACACCATGCTCTGCTATAGATCTGTTTCATGTTTGGTGGGCTGCTGT 4926
Db 2322 ATGAGGAACCTCCATGATATAGATATTGACATGTTTCAATTTATGTTGGGCTTGTGTC 2381
QY 4927 TAGCATTCCTCCTCTGATTTGGATGACCTGTTGATCTCGACCTTCTTCTCAGTTAGTT 4986
Db 2382 TTGCAATTTCTCGTTGCTGCTG-----TCAGGATTTTCAGGGATCAGCC 2426
QY 4987 CTTCTATAACCACTTATCTCTCCATTTGATCACCATGGACACATGCTTCCAGATAC 5046
Db 2427 TTGGCACTGGAGACCTTCACATTTTCCATCTGTTTACTATGGCAGACATCATACAGATCT 2486
QY 5047 TACTTACAGTAGACAC---AGGCCATCCCTTGTCTAGGTTTCAAGAGACAGTGAAGAGG 5103
Db 2487 TACTTACCTCATGTACAGAGAAATGGCATGGATCAAGAAATATCCCTTTGTGAAGAAG 2546
QY 5104 CTCATTCCGATCTCTCTTTTCAGAAATTTCTCAATATACAAAGTGGCTTCATGGGT 5163
Db 2547 AATCAGCAGTTCTTCTGTTGTATTAACACCTTCACCAGTATACGGGAAGTGCCTTGAAG 2606
QY 5164 GTGATATTCCTGGCTGTTTGGGTCTCCTCAGTGAAGATGGCATACCCCTTATCTTC 5223
Db 2607 AAATACCATCGGCTGGCATCTGTGGAGGAGTGTGAGAGTGAATCATGCTTCTCTGA 2666
QY 5224 GCTGTGCTGATGTTTTTTCACATATTTTACTTGGGGTAACTCCGCTGAGGAAGTGCATA 5283

Db 2667 AGTGTCTGCTTATTTTTCATTTAAATGGAGTTCTCTTCCACCCAGACATTTCAAG 2726
QY 5284 CCAATTTCTGCAGAGAGAGTACAGTGCAGCTCTGTAGCTTATCTATCTTTTACCTACAAAT 5343
Db 2727 T---TCCCTGGAACAGCCATTTTGAACATTTATGATCTATCTTCCCTTACCAACAACC 2783
QY 5344 TGTTCCTCTCTCTTCCAGGAATATGGGATCTGTAAAGCCCTTGTCTCCAGAGGTGGTGTG 5403
Db 2784 TCATTTGCTCTTTTCAAGAAATAGTAGATTAATGAATTTCACTGATGAAAGTTGGTGCC 2843
QY 5404 CAGATCCTGCTTACTAACTGTTTGAACCAAAAACACCCGTGCTCAGGTACCTTAGAA 5463
Db 2844 GTAACAGTGAAGTTTAAAGATATCTAGAAGTTGAAGAGATGCTATAAGATATCCAAGAG 2903
QY 5464 AAAAGAAATAGTTTGTATAGAGTTCCTGATGACTATAGCTGCTCTGATTAAGCTTTCTC 5523
Db 2904 AATCTAACAAATTAATAAACCTTCCAGAGGATTACAGACGCTCATTAATCAAGCATCCA 2963
QY 5524 ATTTCAAGTGGCCCGCTCTGCAGATGATGAGCGAAAGCATCTGTCTCTGCTTCTTCT 5583
Db 2964 ATTTCTGTCGCCGAAATCAGGTGTTGATAAGAGCAGAGCCCAACTCTGTGCTTGTGT 3023
QY 5584 GTGGGCTATATCTATGTTCTCAGAACATTTGCTGCCAGGAATTTGTGAACGGGAAGAGG 5643
Db 3024 GCGGATCTCTGCTGTCTCCAGAGTTACTGCTGCCAGACTGAATGGAAGGGGAGGATG 3083
QY 5644 TTGAGCTTTGCAATTTTTCACGCACATCTACTGTGGAGCCGAGTCTGCAATTTTCTCTCA 5703
Db 3084 TAGGAGCTGCACACCTCACACCTACTCTCTGTGCTGTGAGTGGGCACTTCTCTGAGAG 3143
QY 5704 TCAGAGATGCGGAGTGTCTGTTGAAGTTAAAGCCAGAGCTGTGCTATCCAGCTC 5763
Db 3144 TACGGGAATGTGAGTGTCTATTTTGTGCTGCAAAACCAAGGCTGTTTTTATCTCTCTC 3203
QY 5764 CTTACTTGGATCAATATGAGAAACAGACCCCTGGCTGAGAGGGCAACCCCTTCATT 5823
Db 3204 CTTACTTGTAGCTATGGGAGACCCGACCGGAGTCTAGACGGGAAATCTCTTACAT 3263
QY 5824 TATCTGCTGAGCGGTATCGGAAGCTCCATTTGCTTGGCAACAACACTGCATTATAGAAG 5883
Db 3264 TATGCAAGAGCGATTCAAGAGATTTCAGAACCTCTGGCACCACACAGTGTACAGAGG 3323
QY 5884 AGATTGCTAGAGCCAGACAGATATCAGATGTTTATTTGGATTCAACTGGCA 5935
Db 3324 AAATTTGGACATGCACAGGAAGCCAAATCAGACACTGTTGGCTTACTGCTGCA 3375

RESULT 5

; US-09-918-995-36721
; Sequence 36721, Application US/0918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 36721
; LENGTH: 505
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(505)
; OTHER INFORMATION: n = A,T,C or G
; US-09-918-995-36721

Query Match 6.9%; Score 438.4; DB 11; Length 505;
Best Local Similarity 99.8%; Pred. No. 5.7e-111;
Matches 439; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4574 GGTATTCTCTTTGCCACAACAATTTATAGAAATGGATTGAAAGTGCACCTGATGAAG 4633
DB 65 GGTATTCTCTTTGCCACAACAATTTATAGAAATGGATTGAAAGTGCACCTGATGAAG 124

QY 4634 GGATCTCGAGTCCCGATGCTGACCTGGAGACACCTGCGCTTTCACCTATCCAGGCAATGA 4693
DB 125 GGATCTCGAGTCCCGATGCTGACCTGGAGACACCTGCGCTTTCACCTATCCAGGCAATGA 184

QY 4694 AAATCTATTGGGAGATGAAGAAACCTCTCTTTGGAGCACTTCAAAATAGCAGCAATAA 4753
DB 185 AAATCTATTGGGAGATGAAGAAACCTCTCTTTGGAGCACTTCAAAATAGCAGCAATAA 244

QY 4754 TGTCTGAAAGCAATTAATGCAAGTTTGCAAGTTGCACAGAGGATTAACCTGCTCAGGTCCT 4813
DB 245 TGTCTGAAAGCAATTAATGCAAGTTTGCAAGTTGCACAGAGGATTAACCTGCTCAGGTCCT 304

QY 4814 GATACGAACATCTGTTCTCTATAGTTCATAGTTCATAGTTCCTTAACATAAAATCAGAAGA 4873
DB 305 GATACGAACATCTGTTCTCTATAGTTCATAGTTCATAGTTCCTTAACATAAAATCAGAAGA 364

QY 4874 TACACCATGCTTCTCTATAGATCTGTTTCATGTTTGGTGGTCTGTTAGCATTT 4933
DB 365 TACACCATGCTTCTCTATAGATCTGTTTCATGTTTGGTGGTCTGTTAGCATTT 424

QY 4934 CCCATCCTTGTATTTGGGATGACCTGTTGATCTGCAGCCTTCTTCAGTTCCTCTA 4993
DB 425 CCCATCCTTGTATTTGGGATGACCTGTTGATCTGCAGCCTTCTTCAGTTCCTCTA 484

QY 4994 TAACCACTTTATCTCTTCC 5013
DB 485 TAACCACTTTATCTCTTAC 504

RESULT 6

US-09-822-849A-532
; Sequence 532, Application US/09822849A
; Patent No. US20020045170A1
; GENERAL INFORMATION:
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fichtel, Kim
; APPLICANT: Agostino, Michael J.
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakar
; APPLICANT: Graham, James R.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6403
; CURRENT APPLICATION NUMBER: US/09/822,849A
; CURRENT FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: 60/195,582
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 598
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 532
; LENGTH: 972
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-849A-532

Query Match 4.0%; Score 250.6; DB 9; Length 972;
Best Local Similarity 57.0%; Pred. No. 1.8e-58;
Matches 479; Conservative 0; Mismatches 359; Indels 3; Gaps 1;

QY 5095 GTGAAGAGGTCATTCGCGATCTCTTCTTTTGAGAAATTTCTCAATATACAGTGGCT 5154
DB 15 GTGAAGAGGTCATTCGCGATCTCTTCTTTTGATATAAACACTTCACCAAGTATACGGGAAGT 74

QY 5155 CCATTGGGTGTGATATTCTTGCTGGTGGTATTGTGGGTCTCCTACGAAGAATGGCATCACCC 5214
DB 75 CCTTGAAGAAATACCATCGGCTGGCATCTGTGGAGGAGTGTGAGAGTGGAAATCATGC 134

QY 5215 CTTATCTTCGCTGTGCTGCAATTTTTCACCTATTTTTCACCTATTTTTCACCTGAGG 5274
DB 135 CTTTCTCTGAAGTGTCTGCTTTTATTTTTCATTTTAAATGGAGTTTCTTCCCAACCCG 194

QY 5275 AACTGCATACCAATTTGTCAGAGAGGAGTACAGTGCACCTCTGTAGTATCTATCTTTAC 5334
DB 195 ACATTTCAAGT---TCCTGGAAACAAGCCATTTTGAACATTTATGTAGTATCTTCCCTAC 251

QY 5335 CTACAAATTTGTTCTGCTCTTCCAGCAATATTGGGATACTGTAAAGCCCTTGTCTCCAGA 5394
DB 252 CAACCAACCTCATTTTCCCTTTTCAAGAAAATAGTGAGATAAATAATTCATCTGATGAAA 311

QY 5395 GGTGGTGTGCAGATCTGCTTACTTAACCTGTTTGAAGCAAAAACACCTGCTCAGGT 5454
DB 312 GTTGGTGGTGTGCAGTGAAGTTTAAAGATATCTAGAAAGTGAAGAGATGCTATAGAT 371

QY 5455 ACCCTAGAAAAAGAAATAGTTTGTATAGAGTTCCTGTATGATCTATAGTGTCTCTGAATC 5514
DB 372 ATCCAAGAGAAATCTAACAATAATAAACCTTCCAGAGGATTTACAGCAGCTCATTAATC 431

QY 5515 AAGCTTCTCATTTTCAGGTGCCCCAGGCTGTCAGATGATGAGCGAAAGCATCTCTCTCT 5574
DB 432 AAGCATCCAAATTTCTGCTGCCGAAATCAGTGTGTAGATGATGAGCAGAGCCCAACTCTGT 491

QY 5575 GCCTTTCTGCTGGGCTATACTATGTTCTCAGAACATTTTCTGCAGGAAATTTCTGAACG 5634
DB 492 GCCTTGTGCGGATCTCTGCTGTCTCCAGAGTTTACTGCTGCCAGCTGAACCTGGAAG 551

QY 5635 GGAAGAGGTTGGAGCTTTCACGCACTTTCACGCTGAGCGCGAGTCTGCAATTT 5694
DB 552 GGAAGGATCTAGGAGCTGCACAGCTCACACCTACTCTCTGCTGAGTGGGCACTCT 611

QY 5695 TCCTAAAAATCAGAAATGCCGAGTGGTCTGTTGAAGTAAAGCCAGAGGCTGTGCTCT 5754
DB 612 TCCTGAGAGTACGGGAATGTGAGTGTCTATTTTAGTGCAAAAACCAAGCTGTCTTTT 671

QY 5755 ATCCAGCTCTTACTTTGGATGAATATGGAGAAACAGACCTTGGCTGAAGAGGGCAACC 5814
DB 672 ATTCTCTCTTACTTACCTTGATGACTATGGGAGACCGACAGGGAATCAGCGGGAATC 731

QY 5815 CCCTTCATTTATCTCTGCTGAGCGGTATCGGAAGCTCCATTTTGGTCTGGCAACAACACTGCA 5874
DB 732 CTTTACATTTATGCAAGAGCGGATTTCAAGAAGATTCAAGAAGCTCTGCGACCAACACAGTG 791

QY 5875 TTATAGAAGATTTGCTAGGAGCCAAAGAGACTAATCAGATGTTTATTTGGATTCAACTGGC 5934
DB 792 TCACAGAGAAATTTGGACATGCACAGGAAGCCCAATCAGACACTGGTTGGCATTTGACTGGC 851

QY 5935 A 5935
DB 852 A 852

RESULT 7

US-10-071-766-7/c
; Sequence 7, Application US/10071766
; Publication No. US20020192678A1
; GENERAL INFORMATION:
; APPLICANT: Huel-Mei Chen
; TITLE OF INVENTION: GENES EXPRESSED IN SENESENCE
; FILE REFERENCE: PA-0043 US
; CURRENT APPLICATION NUMBER: US/10/071,766
; CURRENT FILING DATE: 2002-02-07
; NUMBER OF SEQ ID NOS: 144
; SOFTWARE: PERL Program
; SEQ ID NO 7
; LENGTH: 4573
; TYPE: DNA
; ORGANISM: Homo sapiens

```

; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020192678A1 1330204.22
; NAME/KEY: unsure
; LOCATION: 3145-3168
; OTHER INFORMATION: a, t, c, g, or other
US-10-071-766-7

Query Match
Best Local Similarity 3.5%; Score 219.8; DB 13; Length 4573;
Matches 441; Conservative 0; Mismatches 307; Indels 17; Gaps 2;

QY 5171 TCCTGCTGGTATTGTGGTCTCACTGAAGAATGCAATCCACCCCTATCTTCGCTGTGC 5230
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4535 TCCGGCTGGCATCTCTGGAGGAGTGTCAAGCTGGAATCATCCCTTCCTGAAGTTTC 4476
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5231 TGCATTGTTTTTCCACTATTTACTTGGGTAACTCCGCTGAGAACTGCATACCAATTC 5290
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4475 TCGTTTATTTTTCATTAATTAAGGAGTTCCTCCACCGACATTCAGT---TCC 4419
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5291 TGCAGAGGAGATACAGTGCATCTGTAGCTATCTTATCTTACCTACAAATTTGTTCCT 5350
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4418 TGAACAAGCCATTTTGAACATTTATGTAGCTATCTTCCCTACCAACAACCTCATTTG 4359
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5351 GCTCTTCCAGGAATATGTGGATCTAAGGCCCTTGTCCAGAGGTGGTGCAGATCC 5410
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4358 CCTTTTCAAGAAATAGTGAATAATGAATTCACGTGAAAGTTGGTGCCTGAACAG 4299
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5411 TGCCTTACTAAACTGTTTGAAGCAAAAACACACCGTGTGCTAGCTACCTAGAAAAGAAA 5470
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4298 TGAAGTTAAAGATATCTAGAAGGTGAAGAGATGCTATAAGATATCCAAGAAATCTAA 4239
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5471 TAGTTTGATAGAGCTTCTGTATGACTATAGCTGCCCTCGAATCAAGCTTCTCATTTTCAG 5530
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4238 CAAATTAATAAACCTTCCAGAGGATTACAGCAGCCTCATTAATCAAGCATCCAATTTCTC 4179
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5531 GTGCCACCGGTGTCAGATGATGAGCGAAAGCATCTGTCTCTGTGCTTTTCTGTGGGGC 5590
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4178 GT-----GGTGATAAGAGCAGAGCCCAACTCTGTGCTTGTGTGCGGATC 4133
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5591 TATACTATGTTTCAGAAACATTTGCTGCCAGGAATTTGTAAGGGGAGAGGTTGGAGC 5650
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4132 TCGCTGTGCTCCAGAGTACTGCTGCCAGACTGAAGGAGGAGGATGTAGGAGC 4073
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5651 TTGCATTTTTCAGCACTTCACTGTGGAGCGGAGTCTGCTATTTTCTCTAAATTCAGAGA 5710
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4072 CTGCACAGCTCACACCTACTCTCTGTGGCTCTGAGTGGGATCTTCTGAGAGTACGGGA 4013
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5711 ATGCCAGTGGTCTCTGGTTGAAGGTAAAGCCAGAGGCTGTGCCTATCCAGCTCCTTACTT 5770
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4012 ATGTCAGGTGCTATTTTGTAGCTGGCAAAACCAAGGCTGTTTTTATCTCTCCTTACCT 3953
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5771 GGATGAATATGGAGAACAGACCCCTGCCCTGAAGAGGGGCAACCCCTTCATTTATCTCG 5830
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3952 TGATGACTATGGGGAGACGACAGGAGCTACAGCGGGAATCCCTTTACATTTATGCAA 3893
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5831 TGAGCGGTATCGGAAGCTCCATTTGTCTGGCAACAACACTGCATTTATATAGAAGATTCG 5890
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3892 AGAGCCATTCAGAAAGATTCAGAAAGTCTGGCACCACACAGTGTACAGAGGAATGG 3833
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5891 TAGGAGCCAGAGACTAATCAGATGTTATTTGGATTCAACTGGCA 5935
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3832 ACATGCACAGGAAGCCAATCAGACACTGGTTGGCAITGACTGGCA 3788
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```

RESULT 8
US-10-027-632-260533
; Sequence 260533, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
```

```

; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 260533
; LENGTH: 578
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-260533

Query Match
Best Local Similarity 3.4%; Score 212.2; DB 13; Length 578;
Matches 235; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 3838 ATAACTCATGTATGACAATACATCAGAAATGCCCTGGGAAAGAGATTCATTATGGAGG 3897
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 226 ATTCACCAAAATATGAAGACAAGTGACCAAAACATGTCTCTCATTTCTGTATCATTTCTT 285
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3898 AAGAGACACCCAGCAGCTAGTCTACTCTAGAAATTTGGTCTCTAAACGGGTC 3957
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 286 TTGCTAGCACCACCAGCAGTCTAGTCTAGAAATTTGGTCTCTAAACGGGTC 345
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3958 CATCTGTTACTGAAAGAGAGTGTGCTGAGTGTGCTTGGCAAGAAGACAGGAGTGA 4017
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 346 CATCTGTTACTGAAAGAGAGTGTGCTGAGTGTGCTTGGCAAGAAGACAGGAGTGA 405
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4018 AAATAGAAAATAATGCCATGCTATATATGCGCTGTGTCAGAAAATCTACTGCTTAACCC 4077
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 406 AAATAGAAAATAATGCCATGCTATATATGCGCTGTGTCAGAAAATCTACTGCTTAACCC 465
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4078 AGCAGAGGGGAAAACCCATAGAACTCTCAGGAG 4110
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 466 AGCAGAGGGGAAAACCCATAGAACTCTCAGGAG 498
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```

RESULT 9
US-09-925-301-490/c
; Sequence 490, Application US/09925301
; Patent No. US2002052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 490
; LENGTH: 1461
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1432)
; OTHER INFORMATION: n equals a,t,g, or c
```

```

; NAME/KEY: misc_feature
; LOCATION: (1452)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-301-490

Query Match      1.7%; Score 105.6; DB 9; Length 1461;
Best Local Similarity 82.7%; Pred. No. 6e-18;
Matches 134; Conservative 0; Mismatches 24; Indels 4; Gaps 1;

QY 123 AAAGGCTTATCATCTGTCTTCCACTTATCCAAAGCTGCTATGGCCACTGCTGTGCCG 182
Db 171 ACAGGCTTATCATCTATCTTCCACTTATCCAAAGCGCTACGGCCACTTCTGTCCCG 112
QY 183 CACCTGGAACAGCGCCAGCCCACTACTGCTTCCACTTACCCTGGTTCTCCACCCCTGA 242
Db 111 CAGTTGGAAGACACACAGCCCGCTACAGCTCCGCTACACCGCTTCTCCGGCGCTGT 52
QY 243 TCAGCTGCTGTGCTGTCGCACTTATCCGCTTCTGCTGCTGTC 284
Db 51 ACAGC----TCCCGCGCCATCTGTGCGGCTCTGCGCTTSC 14

RESULT 10
US-09-960-352-14952/c
; Sequence 14952, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathalagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 14952
; LENGTH: 326
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 64-LIB3057-001-Q1-K1-H8
US-09-960-352-14952

Query Match      1.4%; Score 89.6; DB 10; Length 326;
Best Local Similarity 81.2%; Pred. No. 5.5e-14;
Matches 104; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 123 AAAGGCTTATCATCTGTCTTCCACTTATCCAAAGCTGCTATGGCCACTGCTGTGCCG 182
Db 142 ACAGGCTTATCGTCAATCTTCCACTTGTCCAGTAAACCGCTAGGGCTACTGCTGTCCCA 83
QY 183 CACCTGGAACACCGCCAGCCCACTACTGCTTCCACTACCCTGGTTCTCCACCCCTGA 242
Db 82 CACCGGAGGACACCGCCAGCCCACTACTGCTTCCGCTCCGCCACCCAGCGGCTCTCCGGCATGA 23
QY 243 TCAGCTGC 250
Db 22 GAAGCTGC 15

RESULT 11
US-09-918-995-27470
; Sequence 27470, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076

```

```

; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows version 3.0
; SEQ ID NO 27470
; LENGTH: 465
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(465)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-27470

Query Match      1.0%; Score 65; DB 11; Length 465;
Best Local Similarity 49.6%; Pred. No. 5.1e-07;
Matches 198; Conservative 0; Mismatches 195; Indels 6; Gaps 1;

QY 4257 TTGTGACTGTGAAAGTGGAGAAATATCTTTGGCCCTCTTTTGCAAAATCTCTGTGCAATACTGTG 4316
DB 69 TATGATGTAGAAACGAGAAATCCCTTTGGCCCTCTCTTTGTGAATGCTTGAGTAATACTGTT 128

QY 4317 ATCCCCATTAATTCCTTTTGCAACCTCAAAAGATAAACAGTCAGAAATGCAGATGCTCTTGCT 4376
DB 129 ATTCC-----TCTGCTGCTCTCTCCAGAAATATTTTAAACAACAGGTTAAATTTTTTCA 182

QY 4377 CAACATTTTGACCTGGCAGCGGTGGATACAGACTGTTCTTGCCCAAGAAATATCAGGTTATAAT 4436
DB 183 GACCAACCAAAATCTGACTCAGTGGATTAGAACCAATATCTCAGCAAAATAAAGCATTTACAG 242

QY 4437 ATAAGACATGCTAAAGGAGAAACCCAAATTCCTATTCTTTCTTTAATCAAGAAATGGAGAT 4496
DB 243 TTCTTTAGGAAGAAGAAAGATGACTCTCTAATAATGCGCTCTACAAAGAAATTCAGAAAATGTG 302

QY 4497 TCTACTTTGGAGTTCCATTCCTCCAGTGTGTTGGGCTTGAGTCTTCGATTAAATATTCA 4556
DB 303 GATGAATTACAGTCCCTGAGGGTTTCAGGCTGATTTTCGTCCTAAGATCCCTTATTCT 362

QY 4557 AATGATCATCAAGAAATGGTTATTCTCTTTGGCCCAACAAATTTATAGAAATGGAATGAAA 4616
DB 363 GAGAGCATAAAGAAATGCTAACGACATTTGGAACTGCTACCTACAAGGTGGGACTAAAG 422

QY 4617 GTCCCACTGATGAAAGGATCCTCGAGTCCCATGCTG 4655
DB 423 GTTCATCCCAATGAAGAGGATCCTCGTGTCCCAATG 461

RESULT 12
US-09-908-975-8907
; Sequence 8907, Application US/09908975
; Publication No. US20030165843A1
; GENERAL INFORMATION:
; APPLICANT: SHOSHAN, Avi
; APPLICANT: WASSERMAN, Alon
; APPLICANT: MINTZ, Eli
; APPLICANT: MINTZ, Liat
; APPLICANT: FAIGLER, Simchon
; TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND
; FILE OF INVENTION: THAT POPULATE A TRANSCRIPTOME
; FILE REFERENCE: 36688-0005
; CURRENT APPLICATION NUMBER: US/09/908,975
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: US 60/287,724
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: US 60/221,607
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 32337
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 8907
; LENGTH: 60
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-908-975-8907

```


Search completed: September 27, 2003, 20:27:06
Job time : 995.772 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 27, 2003, 00:42:14 ; Search time 7897.46 Seconds

(without alignments)
19412.893 Million cell updates/sec

Title: US-09-724-126A-1

Perfect score: 6308

Sequence: 1 gccagaattcgccagcaggg.....aatttgatttggtgtttt 6308

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Watch 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vit:*
21: em_gss_fun:*
22: em_gss_nam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2096	33.2	2561	11 AK089616	AK089616 Mus muscu
2	826.6	13.1	960	12 BM472160	BM472160 AGENCOURT
3	806.4	12.8	3641	11 BC044903	BC044903 Mus muscu
4	771.4	12.2	865	13 BQ233617	BQ233617 AGENCOURT

5	740.2	11.7	3058	11 AK078173	AK078173 Mus muscu
6	710.4	11.3	712	13 BX104087	BX104087 Mus muscu
7	687.8	10.9	3627	11 AK083320	AK083320 Mus muscu
8	674.6	10.7	682	9 AI929033	AI929033 aue64c10.y
9	668.8	10.6	898	13 BU155953	BU155953 AGENCOURT
10	664.8	10.5	668	13 BU618516	BU618516 UT-H-FH-
11	656.6	10.4	756	10 BG534574	BG534574 602533425
12	637.4	10.1	641	9 AI361043	AI361043 qy03f11.x
13	635.2	10.1	774	13 BU955331	BU955331 AGENCOURT
14	627.2	9.9	797	12 BG862813	BG862813 602799074
15	594.8	9.4	640	14 CD368666	CD368666 UT-H-F1-
16	585.2	9.3	695	14 BY756516	BY756516 BY756516
17	583.6	9.3	703	12 BM114217	BM114217 L0801D04-
18	576	9.1	729	13 BU703363	BU703363 UT-M-FD0-
19	573.2	9.1	782	12 BU086469	BU086469 602849734
20	560	8.9	560	14 CB159971	CB159971 K-EST0219
21	557	8.8	565	10 BF063405	BF063405 7h89406.x
22	545.4	8.6	684	14 BY764809	BY764809 BY764809
23	529.2	8.4	691	14 BY736138	BY736138 BY736138
24	521.2	8.3	528	13 BX119443	BX119443 BX119443
25	517.6	8.2	659	14 BY750548	BY750548 BY750548
26	512.2	8.1	606	10 BE589438	BE589438 195602 BA
27	511.4	8.1	752	14 CA313244	CA313244 UT-CF-FNO
28	509.4	8.1	633	10 BB622408	BB622408 BB622408
29	501.4	7.9	506	2 HSN07106	Bx486919 Homo sapi
30	499.6	7.9	777	14 CB232501	CB232501 AGENCOURT
31	497.4	7.9	719	13 BU623096	BU623096 UT-H-F11-
32	496	7.9	496	13 BX281065	BX281065 BX281065
33	495.2	7.9	583	12 BI540789	BI540789 454188 MA
34	494	7.8	495	14 CB129142	CB129142 K-EST0178
35	487.2	7.7	1105	13 BU751567	BU751567 CH54001-A
36	479.8	7.6	624	10 BG219270	BG219270 RST9023
37	477.8	7.6	481	9 AI192195	AI192195 qc92e08.x
38	473.6	7.5	609	29 CC200266	CC200266 RRC195 Ba
39	473.4	7.5	583	10 AW971391	AW971391 EST383480
40	471.6	7.5	478	9 AA401319	AA401319 zu63d04.i
41	469.4	7.4	473	13 BX108567	BX108567 BX108567
42	464.4	7.4	584	14 CB35201	CB35201 768595 MA
43	464.4	7.4	830	14 CB588645	CB588645 AGENCOURT
44	454	7.2	522	9 AW291190	AW291190 UT-H-BI2-
45	453.8	7.2	737	13 BU461940	BU461940 603775907

ALIGNMENTS

RESULT 1
AK089616
LOCUS
DEFINITION
2561 bp mRNA linear HTC 05-DEC-2002
Mus musculus activated spleen cDNA, RIKEN full-length enriched library, clone: F830005C07 product: ubiquitin protein ligase E3 component n-recogin 1, full insert sequence.
ACCESSION
AK089616
VERSION
AK089616.1 GI:26354612
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE
1
Carninci, P. and Hayashizaki, Y.
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
PUBMED
11042159
REFERENCE
2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
PUBMED
11042159

REFERENCE
AUTHORS

3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multipillarary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL
MEDLINE
PUBMED

20530913
11076861

REFERENCE
AUTHORS

4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I.,
Saito, T., Okazaki, Y., Gojohori, T., Bono, H., Kasukawa, T., Saito, K.,
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
Kuehl, P., Lewis, S., Matsuo, Y., Nikaudo, I., Pesole, G.,
Quackenbush, J., Schriml, L. M., Staubli, F., Suzuki, R., Tomita, M.,
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,
Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C.,
Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P.,
Marchionni, L., Washima, J., Mazzarelli, J., Mombaerts, P., Nordone, P.,
Rind, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H.,
Toyo-oka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L.,
Wyshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S.
and Hayashizaki, Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)
21085660
11217851

REFERENCE
AUTHORS

5 The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

JOURNAL
MEDLINE
PUBMED

6 (bases 1 to 2561)

REFERENCE
AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
Direct Submission

JOURNAL

Submitted (16-Apr-2002) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)

COMMENT

cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Tissues were provided by Dr. John Todd (Dept. of Medical Genetics
Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome
Trust/MRC Building Addenbrookes Hospital Cambridge) whose
assistance we gratefully acknowledge.

Please visit our web site for further details.

URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/

FEATURES
Source

Location/Qualifiers
1. .2561
/organism="Mus musculus"
/mol_type="mRNA"
/strain="NOD"
/db_xref="FANTOM_DB:F830005C07"
/db_xref="taxon:10090"
/clone="F830005C07"
/tissue_type="activated spleen"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
14. .>2560
/note="unnamed protein product; putative
ubiquitin protein ligase E3 component n-recognin 1
(MGDI|GI:1277977, GB|NM_009461, evidence: BLASTN, 100%,
match=2560)"
/codon_start=1
/protein_id="BAC40933.1"
/db_xref="GI:26354613"
/translation="MADEMDGAERMDVSPPEPLAPQRPASWDDQVDFYATFLHLA
QLVPEIYFEMDPDLEKQESVQMSILTPLEWYLFCEDDPDICLEKLKHSAGFOLCKGV
FKSGETYSRCDADPTCVLCMDCFQSVHKNRHKMTSTGGCCDCGDEKPAWKTG
PFCVDPEPRAGRTKESLCHPLNEEYIAQALDCELAEQHLHTAIDKEERRAVKAGVY
IREKNERYCVLFNDHSHSYDHVYISLQALDCELAEQHLHTAIDKEERRAVKAGVY
ATCOAKEDIKSHSENVSOHLHVEVLSVMAHOKFALRLGSMNKMIMSVDFFQI
FCOACIVPEPSGENSECLISRLMLDAKLYKGAPKILHELIFSEFFEMEMEKLFAMEF
VITYQLQKEYISDDHERSISYALSYQMLVTPLARHLEEQNVISVITETILEVLP
EYLDNRNFKFGYSODKIRGVYVCDLKYILISRPVWTERLQYKQVLEGGFSLFKI
LTCMGMEIERQGVGHIEVDPDEAAIAQOMQLKNILMFQEWACDEDLVLYKE
CHKVMRCSTNFMSSTVQVQLCGHSLSTKYKVEDLVSLHPLSLTLAGLHVRLSR
LGAISLHEFVPEPDSQVEVLPCLVLAQVVAEMRRNGLSLISOVFYQDVK
CREMYDKDILMLQIGASIMDPNKKFLVLYORVELTDAFNKLTSTKDDDLIKOYNTLI
EMQLQVLIIVGERYPVGVNTRREVIMRETHILICIEPMPHSATARNLPENNET
GLENVINKVATFKPGVSGHGVYELKDESLKDFNMYFYHYSKTOHSAEHRM"

BASE COUNT 763 a 509 c 617 g 672 t
ORIGIN

Query Match 33.2%; Score 2096; DB 11; Length 2561;
Best Local Similarity 88.7%; Pred. No. 0;
Matches 2270; Conservative 0; Mismatches 290; Indels 0; Gaps 0;
QY 684 GTTTCCTCAAGATGGCGGACGAGGCGTGGAGGTACTGAGAGGATGGAATCAGCGCG 743
DB 2 GTTTCCTCAAGATGGCGGACGAGGATGGCGGCGCCGCGGAGGATGGAGTTCAGCGCG 61
QY 744 GAGTTACCCAGACCCCTCAGCTCTGGCATCTTGTGGGATCAGCAAGTTGATTTTAT 803
DB 62 GAGCCTCCCTGCGCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 121
QY 804 ACTGCTTTCTTGGCATCTTGGGACAAATTTGTCGCCAGAAATTTACTTTGCTGAATGGAC 863
DB 122 ACTGCTTTCTTACATCTTGGGACAAATTTGTCGCCAGAAATTTACTTTGCTGAATGGAC 181
QY 864 CCAGACTTGGAAAGCAGGAGGAGAAAGTGTAACAATGTCAATATTTCACTCCACTCGAATGG 923
DB 182 CCAGATTGGAAAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 241
QY 924 TACTTATTGGAGAGATCCAGATATTTGCTTGGAGAAATTTGAAGCAGTGGAGCATTT 983
DB 242 TACTTATTGGAGAGATCCAGATATTTGCTTGGAGAAATTTGAAGCAGTGGAGCATTT 301
QY 984 CAGCTTTGGGAGGGTTTCAAAAAGTGGAGAGACACCACTTATTTCTTCAGGGGATTTGCA 1043
DB 302 CAGTGTGTGGGAGGGTTTCAAAAAGTGGAGAGACACCACTTATTTCTTCAGGGGATTTGCA 361
QY 1044 ATTGATCCACATGTGTACTCTGTATGAGTCTGCTCCAGGACAGTGTTCATAAAATCAT 1103
DB 362 ATTGATCCACATGTGTACTCTGTATGAGTCTGCTCCAGGACAGTGTTCATAAAATCAT 421
QY 1104 CGTTCAAGATGATCTTCTTACTTGGAGGAGGGTCTGTGACTCTGTGACTCTGGAGAGACGCA 1163

Db 422 CGTTACAAGATGATCTTCTACTTGGAGGGGCTTCTGTGACGTGGAGACACAGAAGCG 481
QY 1164 TGAAGAACTGGCCCTTTTGTGTAAATCATGAACCTGGAAGAGCAGGTACTATATAAGAG 1223
Db 482 TGAAGAACTGGCCCTTTTGTGTGATCACAGCCTGGAAGAGCAGGTACTATACAAAGAG 541
QY 1224 AATTCAGCGTGTCCGTTGAATGAAGAGGTAAATGTCCAAGCCAGGAAATATTTCCCTTCA 1283
Db 542 AGCTTACATTTGCCCATTTGAATGAAGAGGTAAATGTCTCAAGCCAGGAGAAATATTCCTTGG 601
QY 1284 GTGATAAAATATGTCTGTAGAAATGACTATATGGAAGAGGAAAGAAAGTCCCTCTGAA 1343
Db 602 GTGATAAATACATTTGTAGAAATGACTATATGGAAGAGGAAAGAAAGTTCCTCTGAA 661
QY 1344 CTCAGATGAAGGAGAGAAATGAAGATACTATTGTCTCTCTTTCAATGTAGTGAACACCAT 1403
Db 662 CTCAGATGAAGGAGAGAAATGAACGATACTATTGTCTCTTTTCAACGATGAGCACCAT 721
QY 1404 TCATATGACCACTCATATACAGCCTTACAAAGAGCTTGTGACTGTGAGCTGCGAGAGGCC 1463
Db 722 TCGTATGATCATGTGATCTACAGTCTGCAGAGAGCTCTAGATTGCGAGCTTGCAGAGGCA 781
QY 1464 CAGTTGCATACCACTGCCATTTGACAAAGAGGGTCTGCGGCTGTAAAGCGGGAGCTTAT 1523
Db 782 CAGCTGCACAGCACTGCCATTCGAAGAGGGTCTGCGGCTGTCAAAGCAGGTGTGTAT 841
QY 1524 GCTGCTTGGCAGGAAGCAAGAGATATATAAGAGTCAATTCAGAAATGTCTCTCAACAT 1583
Db 842 GCACTTGGCAGGAAGCAAGAGGATATATAAGAGTCACTCAGAGACGCTCTCAGCAC 901
QY 1584 CCATTTTCATAGAGATTTATCACTCAGAGATTATGGCTCATCAGAAATTTGCTTTCCGT 1643
Db 902 CCCTCCATGTGAAGTGTCTGCACTCGCTCGTGTATGGCTCACCAGAAATTCGCTCTGGC 961
QY 1644 CTTGGTTCCTGGATGAACAAATATGAGCTATTCAAGTCACTTTAGCGAGATCTTTGC 1703
Db 962 CTTGGCTCTGGATGAACAAATATGAGCTATTCAAGTCACTTTAGCAGATATTTGC 1021
QY 1704 CAAGCATGCTTGAAGAAACCTGACTCGGAGAAATCCCTGTCTCATAGCAGGTAAATG 1763
Db 1022 CAGGCTCGCTGAGAGAACCTGGCTCTGAATAATCCCTGTCTTATAGCAGACTAATG 1081
QY 1764 CTTTGGATGCAAGCTTTTATAAGGTGCCCGTAAGATCCTTCATGAATTTGATCTTCAGC 1823
Db 1082 CTTTGGATGCAAACTTTTATAAGGTGCCCGTAAGATCCTTCATGAATTTGATCTTTAGT 1141
QY 1824 AGTTTTTTTATGAGATGGAATACAAAACCTCTTTGCTATGAAATTTGTGAAGTATAT 1883
Db 1142 AGTTTTTTTATGAGATGGAATACAAAACCTCTTTGCTATGAAATTTGTGAAGTATAT 1201
QY 1884 AAACAACTGCAGAAAGAAATATATCAGTGTATCATGACAGAAAGTATCTCTATAACTGCA 1943
Db 1202 AAACAACTGCAGAAAGAAATATATCAGGACGACACGAGAGAGCATCTCCATTAACCGCC 1261
QY 1944 CTTTCAGTTCAGATGTTTACTGTCTCTACTCTGGCTCGACATCTTATTGAAGAGCAAAAT 2003
Db 1262 CTGTCCGTGCAGATGCTCACCGTCCCGACCTTGGCCGGCATCTTATTGAAGAGCAAAAT 1321
QY 2004 GTTATCTCTGATTAATCACTGAACTCTGTAGAGTCTTTACTGATGAAATTTGTGAGAGAAC 2063
Db 1322 GTTATTTCTGCTAATTAAGAGCTGTAGAGTCTTTACTGAAATTTGTGAGAGAAC 1381
QY 2064 AATAAATTAACCTCCAGGTTTATAGCAGGACAAATTTGGAGAGATATATGAGTAATA 2123
Db 1382 AATAAATTAACCTCCAGGTTTATAGCAGGACAAATTTGGAGAGATCTTACGAGTTATA 1441
QY 2124 TGTGACTTAAAGTATATCTGTATCAGCAAAACCAATATGAGCAGAAAGATTAAAGAAATG 2183
Db 1442 TGTGACTTAAAGTATATCTGTATCAGCAAAACCAATATGAGCAGAAAGATTAAAGAGCG 1501
QY 2184 CAGTTCCCTTGAAGGTTTTCGATCTTTTGAAGATCTTTTACCTGTATGCGAGGAATGAA 2243
Db 1502 CAGTTCCCTTGAAGGTTTTCGATCTTTTGAAGATCTTTTACCTGTATGCGAGGAATGAA 1561

RESULT 2
BM472160
LOCUS

BM472160

960 bp

mrna

linear

EST 05-FEB-2002

QY 2244 GAAATCCGAAGACAGGTTGGCAACACATTTGAAGTGGATCCTGATTTGGAGGCTGCCATT 2303
Db 1562 GAAATCAGAAGACAAAGTTGGCAACACACATTTGAAGTGGACCTGCTACTGGAGGCTGCCATC 1621
QY 2304 GCTATACAGATGCAATTTGAAGAAATATTTTACTCATGTTCCAAAGAGTGGTGTGCTGTGAT 2363
Db 1622 GCTATACAGATGCAATTAAGAAATATTTTCTCATGTTTCCAAAGAGTGGTGTGCTGTGAT 1681
QY 2364 GAAGAACTCTTACTTGTGGCTTATAAAGAAATGTCACAAAGCTGTGATGAGGTGCAGTACC 2423
Db 1682 GAAGATCTTCTACTGCTGCTTATAAAGAAATGTCACAAAGCTGTGATGAGGTGCAGTACA 1741
QY 2424 AGTTTTCATATCTAGTAGCAAGACAGTAGTACAAATCGTGTGACATAGTTTGGAAACAAAG 2483
Db 1742 AATTTTCATGTCAGTACCAAGACAGTAGTACAAATCGTGTGACATAGTTTGGAAACAAAG 1801
QY 2484 TCCTACAGATATCTGAGGATCTTGTAGCAGATACATCTGCCACTCTCTAGGACCTTGTCT 2543
Db 1802 TCCTACAAAGTCTCTGAGGACCTTGTAGCAGATACACCTGCCACTCTCTAGAACACTTGTCT 1861
QY 2544 GGTCTTCATGTACGTTTAAAGCAGGCTGGGTGCTGTTTCAAGACTGCATGAATTTGTGTCT 2603
Db 1862 GGTCTTCATGTACGTTTAAAGCAGGCTGGGTGCTGTTTCAAGACTGCATGAATTTGTGTCT 1921
QY 2604 TTTGAGGACTTTTCAAGTAGAGTACTAGTGGAAATATCTTTTACGTTTGTCTGTTGTGTT 2663
Db 1922 TTTGACAGCTTTCAAGTAGAGGCTCTGGTGGAGTACCCGCTGGCTGCTGCTGCTGGT 1981
QY 2664 GCCAGGTTGTTGCTGAGATGCGGAGAGAAATGAGACTGTCTCTTAAATGACCAAGTGT 2723
Db 1982 GCTCAGGTTGTTGCTGAGATGCGGAGAGAAACGGGCTCTCACTCATCAGCCAGGTTTTC 2041
QY 2724 TATTACCAAGATGTTAAGTGCAGAGAGAAATGTATGATAAAGATATCATCGCTTCAG 2783
Db 2042 TATTATCAAGATGTTAAGTGCAGAGAGAAATGTATGATAAAGATATCATCGCTTCAG 2101
QY 2784 ATTGTGCTGCTTTAATGATCCCAATAGTCTTGTACTGTGACTTCCAGAGGATGAA 2843
Db 2102 ATTGAGCATCTATAATGATCCCAACAAAGTCTTGTACTGTCTTCCAGAGATGAA 2161
QY 2844 CTTGCCGAGGCTTTTAAACAGACCATATCTACAAAGACCAAGATTTGATTTAAACAATAT 2903
Db 2162 CTTACTGATGCTTTTAAACAGACCATATCTACAAAGACCAAGATTTGATTTAAACAATAT 2221
QY 2904 AATCACTAATAGAGAAATGCTTCAAGTCTCATCTATATGCTGGTGAGCGCTTATGTA 2963
Db 2222 AATCACTAATAGAGAAATGCTTCAAGTCTCATCTATATGCTGGTGAGCGCTTATGTA 2281
QY 2964 CTTGAGTGGGAAATGTGACCAAGAGAGGTCAAAATGAGAGAAATCATCTACTTGTCT 3023
Db 2282 CTTGAGTGGGAAATGTGACCAAGAGAGGTCAAAATGAGAGAAATCATCTACTTACTT 2341
QY 3024 TGCATTGAACCCATGCCACAGTGCCTATTCGCAAAATTTACCTGAGAAATGAAATAAT 3083
Db 2342 TGCATTGAACCCATGCCACAGTGCCTATTCGCAAAATTTACCTGAGAAATGAAATAAT 2401
QY 3084 GAACTGCTTAGAGAAATGTATATAAAGAGTGGCCACATTTAAGAAACCAAGGTGTATCA 3143
Db 2402 GAACTGCTTAGAGAAATGTATATAAAGAGTGGCCACATTTAAGAAACCAAGGTGTATCA 2461
QY 3144 GGCCATGGAGTTTATGAATCAAAAGTCAAAAGTCAAAAGTCAAAAGTCAAAAGTCAAAAGT 3203
Db 2462 GGCCATGGAGTTTATGAATGAAAGTCAAAAGTCAAAAGTCAAAAGTCAAAAGTCAAAAGT 2521
QY 3204 CATTACTCCAAACCCAGCATAGCAGGCTGAACATATGC 3243
Db 2522 CATTATTCTAAACACACACATAGCAAGGCTGAACATATGC 2561

DEFINITION AGENCOURT_6469668 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:5576383
5', mRNA sequence.
ACCESSION BM472160
VERSION BM472160.1 GI:18521202
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC <http://mgc.nci.nih.gov/>.
NATIONAL INSTITUTES OF HEALTH, MAMMALIAN GENE COLLECTION (MGC)
UNPUBLISHED
CONTACT: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L1AM12328 row: c column: 08
High quality sequence stop: 720.
Location/Qualifiers
1..960
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5576383"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_92"
/note="Organ: testis; Vector: pCMV-SPORT6; Site: 1: NotI;
Site: 2: SalI; Cloned unidirectionally; oligo-dr primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 230 a 227 c 218 g 283 t 2 others
ORIGIN
Query Match 13.1%; Score 826.6; DB 12; Length 960;
Best Local Similarity 96.1%; Pred. No. 7.8e-160;
Matches 890; Conservative 0; Mismatches 30; Indels 6; Gaps 4;
Qy 4924 TGTAGCATCCCATCCCTGTTGATGGGATGACCCCTGTTGATCGAGCCTTCTTCAGTTA 4983
Db 1 TGTAGCATCCCATCCCTGTTGATGGGATGACCCCTGTTGATCGAGCCTTCTTCAGTTA 60
Qy 4984 GTTCTTCTATACACACCTTTATCTCTTCCATTTGATCACCATGGCACACATGCTTCAGA 5043
Db 61 GTTCTTCTATACACACCTTTATCTCTTCCATTTGATCACCATGGCACACATGCTTCAGA 120
Qy 5044 TACTACTACAGTAGACACAGCGCTACCCCTGCTCAGGTTCAAGAAGACAGTGAAGAGG 5103
Db 121 TACTACTACAGTAGACACAGCGCTACCCCTGCTCAGGTTCAAGAAGACAGTGAAGAGG 180
Qy 5104 CTCATTCGGCATCTCTTTCTTTGAGAAATTTCTCAATATACAGTGGCTCCATGGGT 5163
Db 181 CTCATTCGGCATCTCTTTCTTTGAGAAATTTCTCAATATACAGTGGCTCCATGGGT 240
Qy 5164 GTGATATTCCTGGCTGTTATTTGGGTCCTCACTGAAGAATGGCATCCACCCCTTATCTTC 5223
Db 241 GTGATATTCCTGGCTGTTATTTGGGTCCTCACTGAAGAATGGCATCCACCCCTTATCTTC 300
Qy 5224 GCTGTGCGCAATGTTTTCACATATTTACTTGGGGTAACCTCCGCTGAGGAACATGCATA 5283
Db 301 GCTGTGCGCAATGTTTTCACATATTTACTTGGGGTAACCTCCGCTGAGGAACATGCATA 360
Qy 5284 CCAATTCGCAAGAGGAGTACAGTGCACCTCTGTAGCTATCTATCTTACCTACCAAT 5343
Db 361 CCAATTCGCAAGAGGAGTACAGTGCACCTCTGTAGCTATCTATCTTACCTACCAAT 420

Qy 5344 TGTTCCTGCTCTTCCAGGAATATTGGGATACCTGTAAAGGCCCTTGCCTCCAGAGGTGGTGTG 5403
Db 421 TGTTCCTGCTCTTCCAGGAATATTGGGATACCTGTAAAGGCCCTTGCCTCCAGAGGTGGTGTG 480
Qy 5404 CAGATCCTGCCCTTACTAAACTGTTTGAAGCAAAAACACCGTGGTCAAGTACCTAGAA 5463
Db 481 CAGATCCTGCCCTTACTAAACTGTTTGAAGCAAAAACACCGTGGTCAAGTACCTAGAA 540
Qy 5464 AAAGAAATAGTTTGATAGAGCTTCCCTGATGACTATAGTGCCTCCCTGAATCAAGCTTCTC 5523
Db 541 AAAGAAATAGTTTGATAGAGCTTCCCTGATGACTATAGTGCCTCCCTGAATCAAGCTTCTC 600
Qy 5524 ATTTCAAGTGCACCGTCTGCAGATGATGAGGAAAGATCTCTGCTCTGCTGCTTTCT 5583
Db 601 ATTTCAAGTGCACCGTCTGCAGATGATGAGGAAAGATCTCTGCTCTGCTGCTTTCT 660
Qy 5584 GTGGGCTATACATGTTCTCAGAACATTTGCTGCCAGGAAATTTGAACGGGGAAGAGG 5643
Db 661 GTGGGCTATACATGTTCTCAGAACATTTGCTGCCAGGAAATTTGAACGGGGAAGAGG 720
Qy 5644 TTGGAGCTTGCATTTTTCACGCACCTTCACTGTGGAGCCGAGTCTGCATTTTCTTCTTCT 5703
Db 721 TTGGAGCTTGCATTTTTCACGCACCTTCACTGTGGAGCCGAGTCTGCATTTTCTTCTTCT 780
Qy 5704 TCAGAGATGCGGAGTGG-TCCCTGGTTGAAGTAAAGCCAGAGGCTGT-CCCTATCCAGC 5761
Db 781 TCAGAGATGCGGAGTGGTCTGTTGAAGTAAAGCCAGAGGCTGTGCTTATCCATTC 840
Qy 5762 T-CCTTACTTGGATGATATGGAGAA--ACAGACCTTGGCTTGAAGGCGGCAACCC 5817
Db 841 TCCCTTACTTGGGATGATATGAAAAACAGAACCTTGGGCTTGAAGGCGGCAACCC 900
Qy 5818 TTCAATTTATCTGCTGAGCGGTATCGG 5843
Db 901 CCCTTTCATTTTCTTCTGAGCGG 926
RESULT 3
BC044903
LOCUS 3641 bp mRNA linear HTC 10-FEB-2003
DEFINITION Mus musculus, Similar to ubiquitin ligase E3 alpha-II, clone
IMAGE:3493115, mRNA.
ACCESSION BC044903
VERSION BC044903.1 GI:28279967
KEYWORDS HTC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Strausberg, R.
Direct Submission
Submitted (23-JAN-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: angbcm.tmc.edu
Gunnaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Lounseged, H.,
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
A.N., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

This clone has the following problem: frame shifted.

source

989 a	899 c	906 g	847 t
-------	-------	-------	-------

DB 743

Qy	4275	GAATATCTTTGGCCCTCTTTTGCAAATCTCTGTGCAATCTGTGATGCCCATTAATTCCTTTG	4339
Db	1867	GAGTTCTCTTCGCCGCTCTGTGAGTGCCTCAGCAACCGGTGATCCCC-----CTGCTG	1920
Qy	4335	CAACCTCAAAAGATAAACACGTGAGATGCGATGCTCTTGCTCAACTTTTGACCCCTGGCA	4394
Db	1921	CTTCTCCCAAGGACATCCTCGGAGAGGTTAAATTTTTCAGACCAACACGATCTGGCA	1980
Qy	4395	CGGTGGATACAGACTGTTCTTGGCCAGAAATATCAGGTTATATAATTAAGACATGCTAAAGGA	4454
Db	1981	CAGTGGAGAGACGATTAACGCGACAAATAAAGGTGGTCCAGATGCTGAGGACAAAGCAC	2040
Qy	4455	GAAAACCCAAATTCCTATTTTCTTTAATCAAGGAATGGAGATTCTACTTTGGAGTTCCAT	4514
Db	2041	AGTGTCTGCACACGCTTCTTCCAGAGGACACAGAAGCCATGAATAATAATACCGATCCCC	2100
Qy	4515	TCCATCTGTAGTTTGGCGTTGAGTCTTCGATTAAATATTCAATAGCATCAAGGAATG	4574
Db	2101	GAAGGCTTCAGGCGCTGATTTTATPCTTAGAACCCCATATTCTGATAGCATAAAGAANA	2160
Qy	4575	GTTATTCTCTTGGCCACCACAATTTATAGAAATTTGGATTGAAAGTGCCACCTGATGAAAG	4634
Db	2161	TTAAGGACATTTGGAACGGCTGCTTACAAGTGGACTGAAGTTTATCCTTAATGAAGGT	2220
Qy	4635	GATCCTCGAGTCCCCATGCTGACCTGGAGCACCTGGCGTTTTCATATPCCAGGCAATTGAA	4694
Db	2221	GACCCCGTGTGCCATCTGTGCTGGGGACCTGTGCATACACCATCCAGAGCATAGAA	2280
Qy	4695	AATCTATTGGAGATGAAGAAACCTCTCTGTTGGAGCACTTCAAAATAGGCAGCATAT	4754
Db	2281	AGAATTTGAGTATGAGGAGAGACCTGTTTTTGGACCTCTGCTTGATAGCTGGACGAC	2340
Qy	4755	GGTCTGAAAGCATTAATTCGAGTTTGCAGTTGCGACAGAGGATTACCTGCTCCTCAGGCTG	4814
Db	2341	TGCTCAGTCTGTTAAACAGGTTTGCAGCAGCACATTTGGACAGTGGCGTTACTTCTGTG	2400
Qy	4815	ATACAGAAACATCTGGTTCGCTCTCTATCAGTTGTTCTTCCCTAACATATAAATCAGAAG	4874
Db	2401	GTACAGGACACTTCTGTAACCTCTTTGTCATCTTGTGCTGCTAGTGACAGCTATGAAG	2460
Qy	4875	ACACATGCGCTCTGCTATAGATCTGTTTCATGTTTTTGGTGGTGTCTGTGTTAGCATTC	4934
Db	2461	CTCCGCTGCGATPAGACATCGACATGTTTTCACCTGTGCTGGGCGCTGGTGTGCTGTT	2520
Qy	4935	CCATCTTGATTGGGATGACCCCTGTTGATCTGCAGCCTCTTTCAGTTAGTTCTTCTCAT	4994
Db	2521	CCAGCTCTGAGT-----TCAGGATTTTTCAGGAACGACCCCTGGCCACT	2565
Qy	4995	AACACCTTTATCTCTCCATTTTATCACCATGGCACATGCTTCAGATACTACTTACA	5054
Db	2566	GGAGACCTGCACATCTTCCACCTGGTTACCATGGCACACATCGTACAGATCTTACTTACC	2625
Qy	5055	GTAGACAC---AGGCCTACCCCTGCTCAGGTTCAAGAAGACAGTGAAGAGGCTCATTC	5111
Db	2626	TCATGTACAGAAGAGATGGCATGGATCAAGAGAAATCCCATCTGGGAAGAGAACCTGGCC	2685
Qy	5112	GCATCTTTCTTTTTCAGAAAAATTTCTCAATATACAAGTGGCTCCATTTGGTGTGATATT	5171
Db	2686	ATTCCTCTTTTGCACAAAACACTTCACCATGATCTGGAAGTGCC--TTGAAGAAGGCCCC	2744
Qy	5172	CCTGGCTGGTATTTGTGGGTCTCACTGAAGAATGGCATCACCCTTATCTTCGCTGTGCT	5231
Db	2745	TCCGGCTGGCACCTGTGGAGGCGCCGGCCGCATCATGCTTTTCCCTCAAGTGTCT	2804
Qy	5232	GCATTTGTTTTCACATATTTACTTTGGGTTAACTCCGGCTCAGGAACATGCATACCAATTCT	5291
Db	2805	GCTTTGTTTTTCCACTATTTAAATGGAGTCCCGGCCCTCCAGACCTTCAAGT---TTCT	2861
Qy	5292	GCAGAAGGAGTACATGTCACCTGTGTAGCTATCTATCTTTTACCTACAAATTTGTTCCCTG	5351
Db	2862	GGACAAGCCATTTTGAACACTTATGTAACTTACCTTTTCCCTTACCACCAACCTCATTTAC	2921
Qy	5352	CTCTTCCAGGAATATTGGGATCTGTAAAGGCCCTTGCTCCAGAGGTTGGTGTGCAGATPCT	5411

Db	2922	CTTTTTCAGAAAAACAGTGCATCATGAACCTCCCTGATTGAAAGTTGGTGCCAGAACAGT	2981	
QY	5412	GCCTTACTAACTGCTTTTGAAGCAAAAAACACCGTGCAGTACCGTACCTAGAAAAGTTCAGTCCAGAAAGAAT	5471	
Db	2982	GAAGTTAAACGGTATCTAAATGGCGAGAGAGCGATAAGCTACCCACGAGGAGCTAAC	3041	
QY	5472	AGTTTGATAGAGCTTCCTGATGACTATAGCTGCCTCCTGAATCAAGCTTCTCATTTTCAGG	5531	
Db	3042	AAACTGATAGACCTTCAGAGGATTACAGCAGCTCATTAACCAAGCATCCAATTTCTCG	3101	
QY	5532	TGCCACGGCTGCAGATGATGAGCGAAAGCATCCTGCTCTGCTGCTTTTCTGTGTGGGGCT	5591	
Db	3102	TGCCCAAAATCAGGTGGCGACAAGACAGCAGAGCTCCTACTCTGTGCTCGTGTGTGGGAGT	3161	
QY	5592	ATACTATGTTCTCAGACATTTGCTGCCAGGAAATTTGACGCGGGGAAGAGTTGGAGCT	5651	
Db	3162	CTCCTCTGCTCTCAGAGTTACTGTCTGCCAAGCTGAGCTGGAGGTGAGGACGTCGGAGCC	3221	
QY	5652	TGCATTTTTCACGCACCTCACTGTGGAGCGCGAGCTGCAATTTTCCCTAAAAATCAGAGAA	5711	
Db	3222	TGCACAGCACACCTACTCCTGTGGCTCCGGGCGCGCATCTTCCTGAGAGTTCGGGAA	3281	
QY	5712	TGCCGAGTGGTCTGGTTGAAGGTAAAGCCAGAGGCTGTGCCTATCCAGCTCCTTACTTG	5771	
Db	3282	TGTCAGGTGCTATTATTTAGCTGGCCAAACCAAGAGATCTTTTATTCTCCTCTTACCTT	3341	
QY	5772	GATCAATATGAGAAACAGACCCCTGGCTGGAAGAGGCGAACCCCTTCATTATATCTCGT	5831	
Db	3342	GACGACTTGGGAGACCGACGAGGACTCAGACGAGAAATCTTCATTATTAAGCAA	3401	
QY	5832	GAGCGGTATCGGAAGCTCCATTTTGGTCTGGCAACACACTGCATTATAGAGAGATTGCT	5891	
Db	3402	GAGCGGTTTCGAAAGATCCAGAACTCTGGCAGCAGCATAGTATCAGAGGAGATCGGA	3461	
QY	5892	AGGAGCCAGAGACATAATCAGATGTTATTTGGATTCAACTGGCAGTTACTGTGAGCTC	5949	
Db	3462	CACGCGCAGGAGGCTAACCCAGACCCCTGGTCGGAATTGACTGGCAGCATTTATATCGC	3519	
RESULT 4	BQ233617	865 bp	linear	EST 02-MAY-2002
LOCUS	AGENCOURT_7575596	NIH_MGC_72	Homo sapiens	cdna clone IMAGE:6047379
DEFINITION	5', mRNA sequence.			
ACCESSION	BQ233617			
VERSION	BQ233617.1	GI:20415017		
KEYWORDS	EST.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
TITLE	1 (bases 1 to 865)			
JOURNAL	NIH-MGC http://imgc.ncbi.nih.gov/.			
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)			
	Unpublished			
	Contact: Robert Strausberg, Ph.D.			
	Email: cgapbs-r@mail.nih.gov			
	Tissue Procurement: ATCC/DCTD/DTF			
	cdna Library Preparation: Life Technologies, Inc.			
	CNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)			
	DNA Sequencing by: Agencourt Bioscience Corporation			
	Clone distribution: MGC clone distribution information can be			
	found through the I.M.A.G.E. Consortium/LLNL at:			
	http://image.llnl.gov			
	Plate: LLNL13294	row: 1	column: 04	
	High quality sequence stop: 720.			
FEATURES	Location/Qualifiers			
source	1..865			
	/organism="Homo sapiens"			
	/mol_type="mRNA"			
	/db_xref="taxon:9606"			
	/clone="IMAGE:6047379"			

```
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_72"
/notes="Organ: Skin; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2 kb. Library constructed by Life
Technologies."
BASE COUNT      261 a 146 c 204 g 250 t 4 others
ORIGIN

Query Match
Best Local Similarity 12.2%; Score 771.4; DB 13; Length 865;
Matches 786; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

QY 2341 TCCAAGAGCGTGTCTTGTGATGAGAACTCTTACTTGTGGCTTATAAAGATGTCACA 2400
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
8 TCCGAGAGTGGTGTCTTGTGATGAGAACTCTTACTTGTGGCTTATAAAGATGTCACA 67

QY 2401 AAGCTGTGATGAGTGCATGACAGTTTTCATATCTAGTAGCAAGACAGTAGTCAATCGT 2460
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
68 AAGCTGTGATGAGTGCATGACAGTTTTCATATCTAGTAGCAAGACAGTAGTCAATCGT 127

QY 2461 GTGGACATAGTTTGGAAACAAAGTCTCTACAGAGTATCTGAGGATCTTGTAAAGCATACATC 2520
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
128 GTGGACATAGTTTGGAAACAAAGTCTCTACAGAGTATCTGAGGATCTTGTAAAGCATACATC 187

QY 2521 TGCCACTCTAGGACCCCTGCTGGTCTTCATGTCAGTTTAAGCAGCGTGGTGCCTGTTT 2580
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
188 TGCCACTCTAGGACCCCTGCTGGTCTTCATGTCAGTTTAAGCAGCGTGGTGCCTGTTT 247

QY 2581 CAAGACTGCATCAATTTGTGCTTTTGGAGGACTTCAAGTAGAGGTAAGTGAATATC 2640
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
248 CAAGACTGCATCAATTTGTGCTTTTGGAGGACTTCAAGTAGAGGTAAGTGAATATC 307

QY 2641 CTTTACGTTGCTGTTGTTGCTGCCCAGGTTGTTGCTGAGAGTGTGCGGAAGAAATGGAC 2700
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
308 CTTTACGTTGCTGTTGTTGCTGCCCAGGTTGTTGCTGAGAGTGTGCGGAAGAAATGGAC 367

QY 2701 TGTCTCTTATGACAGGTTGTTTATACCAAGATGTTAAGTCGACAGAGAAATGTATG 2760
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
368 TGTCTCTTATGACAGGTTGTTTATACCAAGATGTTAAGTCGACAGAGAAATGTATG 427

QY 2761 ATAAAGATATCATGCTTCAGATGTTGGTGCATCTTTATGATGCCATCCATAGTCTTGT 2820
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
428 ATAAAGATATCATGCTTCAGATGTTGGTGCATCTTTATGATGCCATCCATAGTCTTGT 487

QY 2821 TACTGCTACTTCAGAGGTATGAACCTTGCAGGCTTTTAAACAGACCATATCTACAAAAG 2880
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
488 TACTGCTACTTCAGAGGTATGAACCTTGCAGGCTTTTAAACAGACCATATCTACAAAAG 547

QY 2881 ACCAGGATTTGATTAACAAATATATACACTAATAGAAGAAATGCTTCAGGTCCTCATCT 2940
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
548 ACCAGGATTTGATTAACAAATATATACACTAATAGAAGAAATGCTTCAGGTCCTCATCT 607

QY 2941 ATATTGTGGTGCAGCTTATGTACCTGGAGTGGGAATGTGACCAAGAGAGGTGCACAA 3000
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
608 ATATTGTGGTGCAGCTTATGTACCTGGAGTGGGAATGTGACCAAGAGAGGTGCACAA 667

QY 3001 TGAGAGAATCATTCATCTGCTTGTGATTGAACCCATGCCACAGTGCCTATTCGCCAAA 3060
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
668 TGAGAGAATCATTCATCTGCTTGTGATTGAACCCATGCCACAGTGCCTATTCGCCAAA 727

QY 3061 ATTTACCTGAGAATGAAATATGAACCT-GGCTTAGAGAATGTCTATAACAAAGTGGCC 3119
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
728 ATTTACCTGAGAATGAAATATGAACCTGGGCTTANAGAATGTCTATAACAAAGTGGCC 787

QY 3120 ACATTTAGAACCC 3133
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
788 ACATTTAGAACCC 801

RESULT 5
AK078173
```

LOCUS AK078173 3058 bp mRNA linear HTC 05-DEC-2002
DEFINITION Mus musculus adult male olfactory brain cDNA, RIKEN full-length
E3 component library, clone:6430402E23 product:ubiquitin protein ligase
AK078173.1 GI:26347022
VERSION HTC; CAP trapper.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus

REFERENCE
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159

REFERENCE
AUTHORS Carninci, P., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Todawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861

REFERENCE
AUTHORS Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yananaka, I.,
Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
Kuehl, P., Lewis, S., Matsuo, Y., Nikaide, I., Pesole, G.,
Quackenbush, J., Schirml, L. M., Staubli, F., Suzuki, R., Tomita, M.,
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,
Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Buit, C.,
Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P.,
Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P.,
Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H.,
Toyooka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L.,
Wyshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S.,
and Hayashizaki, Y.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409 (6821), 685-690 (2001)
MEDLINE 21085660
PUBMED 11217851

REFERENCE
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
MEDLINE 12117851
PUBMED 12117851

REFERENCE
AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashizaki, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,

Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

Direct Submission
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.c.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/
Location/Qualifiers
1. 3038
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM.DB:6430402E23"
/db_xref="taxon:10090"
/clone="6430402E23"
/sex="male"
/tissue_type="olfactory brain"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
<1. 691
/note="unnamed protein product: putative ubiquitin protein ligase E3 component n-recognin 1 (MGJ|MGJ:1277977, GI|NM_009461, evidence: BLASTN, 100%, match=2560)"
/codon_start=2
/protein_id="BAC37160.1"
/db_xref="GI:26347023"
/translation="GVAPPEELFANSAGEBFSALCSYLSPTNLFLLFOEYWDTRIRPL LORWCGDPAALLKSLKOKSAVVYPRKRNSLIEPDIYSCILNQASHPRSPSADDERK HPVLCFPGAILCSQICQEIIVNGEYVACVFAHLCGAGVCIFLKIKECRVIVLEG KARGCATPAPYLDYEGTDFELKRGNPLHLSRERYRKLHLVWQOHCIIIEARSQETN QMLFGFNWQLL"
/note="putative"
3032. 3037
/note="putative"
3058
/note="putative"
BASE COUNT 870 a 562 c 632 g 994 t

polyA_signal
polyA_site
BASE COUNT 870 a 562 c 632 g 994 t
ORIGIN
Query Match 11.7%; Score 740.2; DB 11; Length 3058;
Best Local Similarity 82.4%; Pred. No. 6.2e-142;
Matches 904; Conservative 0; Mismatches 148; Indels 45; Gaps 3;
Qy 5255 TGGGGTAACTCCGCTGAGGAACATGACATACCAATCTTCGAGAAGGAGAGTAGTACAGTGCAC 5314
Dy 1 TGGAGTAGCTCCGCTGAGAACTGTTTGCCAAATCTGCTGAAGGAGAAATTCAGTGCAC 60
Qy 5315 CTGTAGCTATCTATCTTTTACATACAAATTTGCTGCTCTTCAGGAATATTTGGGATPAC 5374
Dy 61 CTGTAGCTATCTATCTTTTACCAACAAATTTGCTGCTCTTCAGGAATATTTGGGATPAC 120
Qy 5375 TGTAAAGCCCTTGTCTCAGAGGTGGTGTGAGATCCTGCCCTTACTAACTGTTTGAAGCA 5434
Dy 121 CATAAAGCCCTTACTCAGAGGTGGTGTGAGATCCTGCCCTTACTAACTGTTTGAAGCA 180
Qy 5435 AAAAAACACCCGTGCTCAGGTACCCCTAGAAAAAGAAATAGTTTGTATGAGCTTCCTGATGA 5494

181 GAAAAGTCTGTGGTCAAGTACCTAGAAAAAGAAATAGTTTGTATAGAGCTTCTCTGAGGA 240
5495 CTATAGCTGCCCTCCCTGAATCAAGCTTCTCATTTCAGGTCCCAAGGTCTGCGAGATGATGA 5554
241 CTACAGCTGTCTTCTTAATCAGGCTTCTCATTTCAGGTCTGCGAGATGATGA 300
5555 GCGAAAGCATCTGTCTCTGCTGCTTCTTCTGTGGGGCTATACTATGTTCTCAGAACATTTG 5614
301 GCGAAAGCATCTGTCTCTGCTGCTTCTTCTGTGGGGCTCCTGCTGTTCTCAGAACATCTG 360
5615 CTGCCAGGAAATTTGAAAGGGGAGAGGTTGAGGCTTCGATTTTTCAGCAGCTTCTCAGT 5674
361 TTGCCAAGAAATAGTGAATGGGAGAGGTTGGAGGCTGCGTTCATGTCGCTTCTCATG 420
5675 TGGAGCCGAGTCTGCATTTTCTTAAATATCAGAAATCCGAGTGGTCTGCTGTTGAAGG 5734
421 TGGTCTGAGCTGCTGCATTTTCTTAAATCCGAGAAATCAGGCTGCTGCTGTTGAAGG 480
5735 TAAAGCCAGAGCTGTGCTATCCAGCTCCTTACTTTGGATGATATGGAGAACAGACCC 5794
481 AAAAGCCAGAGCTGTGCTTACCCAGCCCTTACTTTGGATGATATGGAGAACAGACCC 540
5795 TGGCCTGAGAGGGGCAACCCCTTCTATCTCTGAGCGGTATCGGAAGCTCCATTT 5854
541 AGGGCTAAGAGAGGAAACCCCTTCTATCTCTGAGCGGTATCGGAAGCTGCTATTT 600
5855 GGCTGGCAACAACTGCTATATAGAGAGATTTGCTAGGAGCAAGAGACTAATCAGAT 5914
601 GGCTGGCAACAGCACTGCATTTATAGAGAGATTTGCTGAGGAGCAAGAGACTAATCAGAT 660
5915 GTATTTGGATTTCAACTGGCAGTTACTGTGAGCTCCAACTTCTGAGGCTGAGGAG 5974
661 GCTATTTGGATTTCAACTGGCAGTTACTGTGAGCTCCAGTTCTGCTCAAGACAATCATGA 720
5975 ATGACAGAGTAGTAAAGCTGATTCAAAATTTATGAAAAACTTCTGAGGCTGAGGAG 6034
721 CTGACATCAATTAATAAGACTGATTAATAATTTCTAGAGAACTTCTGAGGAG -GGGGAAG 779
6035 TATTCGAGGCTCTTTTGGTCCATGTCAGGTTTCACTTACATCAATAAATAATTTCTTAAT 6094
780 TATTCGAGGCTCTTTTGGTCCATGTCAGGTTTCACTTACATCAATAAATAATTTCTTAAT 839
6095 GGAGTATTCGTTTCAATTAGCAACATATGCTTCAAGGAAAAA -AGGACATAGATCAAT 6153
840 GGAATATTCGTTTCAATTAGCAACATATGCTTCAAGGAAAAA -AGGACATAGATCAAT 899
6154 CTGT-----TTTATGCTAGT 6170
900 GTTTTATGTTTCAAGCACTAAAGAAATGCTTTTTCATCCAGTGTCTATTTCTGCTAAT 959
6171 ATTTCCAGGAATTTATTTCCCTTCATAAATTTGCTCATTTTTCATTTTATTTCCACTTG 6230
960 ATTTCCAGGAATTTATTTCCCTTCATAAATTTGCTCATTTTTCATTTTATTTCCACTTG 1019
6231 GTAGATGAGTCAAGCTCAACAGAGTTGTGAGACATTTTATGTTGGTAACTTCTTCTGCAA 6290
1020 GTTATGAGTCAATTAAGCATTTGTGACATTTTCCATCTCCATCTGCGCTAACATCTCTGCAC 1079
6291 TTTTGTATTTGTTT 6307
1080 CTTTGTATTTGTTT 1096
RESULT 6
BX104087
LOCUS
DEFINITION
BX104087 Soares_testis_NHT Homo sapiens cDNA clone linear EST 06-FEB-2003
; IMAGE:742663, mRNA sequence.
ACCESSION
BX104087
VERSION
BX104087.1 GI:27846028
KEYWORDS
EST.
Homo sapiens (human)

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 712)
REFERENCE Ebert,L., Heil,O., Hennig,S., Neubert,P., Partsch,E., Peters,M.,
Radelof,U., Schneider,D. and Korn,B.
AUTHORS Human Unigeneset - RZPD3
TITLE Unpublished
JOURNAL
COMMENT Contact: Ina Rolfes
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany
RZPD; IMAGp998F081824.
RZPDLIB; I.M.A.G.E. cDNA Clone Collection;
Human Unigeneset - RZPD3 (RZPDLIB No.972)
http://www.rzpd.de/CloneCards/cgi-
bin/showlib.pl.cgi?response=libNo=972 Contact: Ina Rolfes
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Heubnerweg 6, D-14059 Berlin, Germany
Tel: +49 30 32639 101
Fax: +49 30 32639 111
www.rzpd.de
This clone is available royalty-free from RZPD;
contact RZPD (clone@rzpd.de) for further information. Seq primer:
M13r. Primer sequence: TTTCACACAGGAAACAGCTATGAC.
FEATURES
Location/Qualifiers
1..712
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGp998F081824 ; IMAGE:742663"
/sex="male"
/lab_host="DH10B"
/clone_lib="Soares_testis_NHT"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech Laboratories
Inc., and primed with a Not I - oligo(dT) primer [5',
TGTTCACCAATCTGAAGTGGAGCGCGCCCAATTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 174 a 164 c 173 g 201 t
ORIGIN
Query Match 11.3%; Score 710.4; DB 13; Length 712;
Best Local Similarity 99.9%; Pred. No. 7.4e-136;
Matches 711; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 5173 CTGGCTGGTATTGTGGGTCTCACTGAAGAATGGCATGCCCTTATCTTCGCTGTGCTG 5232
Db 1 CTGGCTGGTATTGTGGGTCTCACTGAAGAATGGCATGCCCTTATCTTCGCTGTGCTG 60
QY 5233 CATTTGTTTCCACTATTTACTTGGGGTAACCTCCGCTGAGGAACATGACCAATCTG 5292
Db 61 CATTTGTTTCCACTATTTACTTGGGGTAACCTCCGCTGAGGAACATGACCAATCTG 120
QY 5293 CAGAAGGAGTACAGTGCACCTCTGAGCTATCTATCTTTACCTACAAATTTGTCCTGC 5352
Db 121 CAGAAGGAGTACAGTGCACCTCTGAGCTATCTATCTTTACCTGCAAAATTTGTCCTGC 180
QY 5353 TCTTCAGGAATATTGGGACTGTGAAGCCCTTGCTCCAGAGGTGGTGTGCAGATCCTG 5412
Db 181 TCTTCAGGAATATTGGGACTGTGAAGCCCTTGCTCCAGAGGTGGTGTGCAGATCCTG 240
QY 5413 CCTTACTAACTGTTTGAAGCAAAAAACACCGCTGGTGCAGGTACCCCTAGAAAAAGAAATA 5472
Db 241 CCTTACTAACTGTTTGAAGCAAAAAACACCGCTGGTGCAGGTACCCCTAGAAAAAGAAATA 300
QY 5473 GTTTGATAGAGCTTCTGTAGTACTATAGTGCCTCTGTGATCAAGCTTCTCATTTGAGGT 5532
Db 301 GTTTGATAGAGCTTCTGTAGTACTATAGTGCCTCTGTGATCAAGCTTCTCATTTGAGGT 360

QY 5533 GCCACAGGTCTGCAGATGATGAGCGAAACATCCTCTCTCTGCCCTTTTCTGTGGGGCTA 5592
Db 361 GCCACAGGTCTGCAGATGATGAGCGAAACATCCTCTCTCTGCCCTTTTCTGTGGGGCTA 420
QY 5593 TACTATGTTCTCAGAACATTTCTGCCAGGAATGTGAACGGGAGAGGTTGGAGCTT 5652
Db 421 TACTATGTTCTCAGAACATTTCTGCCAGGAATGTGAACGGGAGAGGTTGGAGCTT 480
QY 5653 GCATTTTTCAGCACTTCACTGTGGAGCGCGAGTCTGCAATTTTCCCTAAAAATCAGAGAAT 5712
Db 481 GCATTTTTCAGCACTTCACTGTGGAGCGCGAGTCTGCAATTTTCCCTAAAAATCAGAGAAT 540
QY 5713 GCCAGTGTCTGTGTTGAAGTAAAGCCAGAGGCTGCGCTATCCAGCTCCTTACTTGG 5772
Db 541 GCCAGTGTCTGTGTTGAAGTAAAGCCAGAGGCTGCGCTATCCAGCTCCTTACTTGG 600
QY 5773 ATGAATATGGAGAACACAGCCCTGCGCTGAACAGGGGCAACCCCTTCAATTCATCCTG 5832
Db 601 ATGAATATGGAGAACACAGCCCTGCGCTGAACAGGGGCAACCCCTTCAATTCATCCTG 660
QY 5833 AGCGTATCGGAAGCTCCATTTGGTCTGGCAACAACACATGCATTATAGAAGA 5884
Db 661 AGCGTATCGGAAGCTCCATTTGGTCTGGCAACAACACATGCATTATAGAAGA 712
RESULT 7
AK083320
LOCUS
DEFINITION Mus musculus 2 days neonate thymus thymic cells cDNA, RIKEN
AK083320 3627 bp mRNA linear HTC 05-DEC-2002
full-length enriched library, clone:C920004H05 product:similar to
(A PROTEIN SIMILAR TO UBIQUITIN-PROTEIN LIGASE E3 COMPONENT
N-RECOGNIN (UBIQUITIN-PROTEIN LIGASE E3-ALPHA)) (FRAGMENT) [Homo
sapiens], full insert sequence.
ACCESSION AK083320
VERSION AK083320.1 GI:26350448
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 Carninci,P. and Hayashizaki,Y.
AUTHORS High-efficiency full-length cDNA cloning
TITLE Meth. Enzymol. 303, 19-44 (1999)
JOURNAL 99279253
MEDLINE
PUBMED 10349636
REFERENCE 2 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Sugahara,K.,
AUTHORS Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
JOURNAL prepare full-length cDNA libraries for rapid discovery of new genes
MEDLINE Genome Res. 10 (10), 1617-1630 (2000)
PUBMED 20499374
REFERENCE 3 RIKEN integrated sequence analysis (RISA) system--384-format
AUTHORS sequencing pipeline with 384 multicapillary sequencer
TITLE Genome Res. 10 (11), 1757-1771 (2000)
JOURNAL 20530913
MEDLINE
PUBMED 11076861
REFERENCE 4 Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
AUTHORS Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
TITLE Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,

Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaado, I., Pesole, G., Quackenbush, J., Schriml, L.M., Staubli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Rinchald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seva, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyooka, K., Wang, K.H., Weitz, C., Whittaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kotsuki, S. and Hayashizaki, Y.

Nature 409 (6821), 685-690 (2001)

21085660

11217851

TITLE
JOURNAL
MEDLINE
PUBMED
AUTHORS

The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

6 (bases 1 to 3627)

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M., Koya, S., Kurahara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Soabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, I., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

Direct Submission

Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan [E-mail: genome-res@gsc.riken.go.jp, URL: <http://genome.gsc.riken.go.jp/>, Tel: 81-45-503-9222, Fax: 81-45-503-9216]

cDNA library was prepared and sequenced in Mouse Genome

Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.

Please visit our web site for further details.

URL: <http://genome.gsc.riken.go.jp/>

URL: <http://fantom.gsc.riken.go.jp/>

Location/Qualifiers

1..3627

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="FANTOM_DB:C920004H05"

/db_xref="taxon:10090"

/clone="C920004H05"

/cell_type="thymic cells"

/tissue_type="thymus"

/clone.lib="RIKEN full-length enriched mouse cDNA library"

/dev_stage="2 days neonate"

1044..>3626

/note="unnamed protein product; putative

similar to (A PROTEIN SIMILAR TO UBIQUITIN-PROTEIN LIGASE

CDS

E3 COMPONENT N-RECOGNIN (UBIQUITIN-PROTEIN LIGASE E3-ALPHA) (FRAGMENT) [Homo sapiens] (SPTR|Q9H578, evidence: FASTY, 78.8%ID, 100%length, match=749), /codon_start=1 /protein_id="BAC38864.1" /db_xref="GI:26350449" /translation="MFRYAVDILTWKESLPEDLEVAEKSDTYCYMLNDEVHTVEQVIYTLQKAVNCTQKEAIGFATTVDRGRSRVYGFQYCDQAKTVIVRNTSRQTKPLKVOVMHSSVAHQNFGLKALSWLGSVIGYSDGLRRLICQVGLQPGDGENSSLDRLMLNDSKMLKWSAGVYHQLFMSLLMDLKYKFLALFRFAKNYERLQDYVTDHDFEVSVALDSVQITVPSLARMLLTENMLTVIIKAFMDHLKRDQAOGRFOFERYTALQAFKFRVLSQILDLKYLKSPTEWDELROKFLQGFDAFLLELLKCMQGMDPITRQVGGHIEMEPWERAFTLOMLKTHVISWVQWCAIDKVLIEAYKKCLAVLTQCHGGFTDGGQPITL SICGHSVETIRYCSQEKVSIHLPLSRLLAGLHVLLSKSEVAFKFPPELLPSLSPPM LLEHLRCLVLCAQVHAGMMRRNFSLVNQIYYHNVKCRREMDKIDVLMQTSVSM LLEHLRCLVLCAQVHAGMMRRNFSLVNQIYYHNVKCRREMDKIDVLMQTSVSM RNPFGVQVAATDEIKREIHLHSIKPMASSELVSKVPDENKENTKESVIESVIEHFKEKPLTGRGMVELKPEKABENLYFYHFSRAEOSKABENOKLREKEDPALPPALP PFCPLPASLVNIIQCDVMLYIMGTILQWAVEHGHSAWSESNLQRYVLIHGMALQEEKH HLENAVEGHVQTFITFTQKISKPGDAPHNSPILAMLETQNASLQNSLEAHKMDIRLLKM FNAIKKIRECSSSPVAEAGTITMESSRDKDKAERKAEIARLRRETKIAQMFEMQ RHFID"

BASE COUNT	943 a	849 c	1012 g	823 t
ORIGIN				
Query Match	10.9%	Score 687.8;	DB 11;	Length 3627;
Best Local Similarity	56.0%;	Pred. No. 4.2e-131;		
Matches 1472;	Conservative	0;	Mismatches 1107;	Indels 51; Gaps 7;
QY	1245	GAAGAGGTAAATGTCCAAAGCCAGGAAAATATATTTTCCTTCAGTGATAAAATATGTCGTAGAA	1304	
DB	1005	GAAGATGTGATGCCAGAACTTACAACATTTTTCGTATTTATTTTCGATATGACGATAGAT	1064	
QY	1305	ATGACTATATGGGAAGAGGAAAGAACTGCCCTCCCTGAACCTCCAGATAGAGGAGAAAT	1364	
DB	1065	ATACTGACCTGGGAAAAGAAAGTGAATGCCCTGAAGACTTAGAAGTGGCAGAGAGAGT	1124	
QY	1365	GAAGATACATATTTGTCCTTTTCAATGATGAACACACATTCATATGACACGTCATATAC	1424	
DB	1125	GACACCTACTACTGCTGCTGTTTAAATGATGAGGTTTACACCTATGACGAAGTCATTTAT	1184	
QY	1425	AGCTTACAAGAGCTCTTGACTGTGAGTCGCGAGAGCCCGCCAGTTCACCTACCTGCGCAT	1484	
DB	1185	ACCTTCGAAAGCTGTGAATGTACACAGAAAGCCATTCGCTTTGCAAGTACAGTT	1244	
QY	1485	GACAAAGAGGTCGTCGGCTGTAAAGCGGAGCTTATGCTGCTGCCAGGAGCAAG	1544	
DB	1245	GATCGAGATGGCGTAGTCTGTCGGATATGGAGATTTCCAGTACTGTGATCAAGCAAG	1304	
QY	1545	GAAGATATAAGAGTCAATTCAGAAAATGCTCTCAACATCCACTTCATGTAGAGTATTA	1604	
DB	1305	ACAGTCATTGTGAGGAACACACAGACAGACAC--AAGCCGCTCAAAGTTCAGTATG	1361	
QY	1605	CACCTCAGAGTATGCTCATCAGAAATTTGCTTGGCTTGGTCTGCTGCTGCTGCTGCTGCT	1664	
DB	1362	CACCTCCTCGCTGCTGCTCATCAGAAATTTGTTGTTGAAAGCTCTGCTGCTGCTGCTGCT	1421	
QY	1665	ATTATCAGCTATTCAAGTGAAGTCTTTAGGCAGATCTTTTGGCAAGCATGCCCTTAGAGAGAA	1724	
DB	1422	GTATTGGATACACTCAGTGGCCTTCCAGAGATTTTGTGTCAAGTGGATTACAAGAGGT	1481	
QY	1725	CCTGACTCGGAGAAATCCCTGTCTCATTAAGCAGAGTTAATGTTGGGATGCAAGCTTTAT	1784	
DB	1482	CCAGATGGCGCAAAACTCTCTCTGTCGACAGACTGATGCTTAATGATTTCAAAATATGG	1541	
QY	1785	AAAGTGGCCGTAAGATCTTTCATGATTCATCTTCAGCAGTTTTTTTTATGGAGATGAA	1844	
DB	1542	AAAGGGCTAGGAGTGTGTATCATCACCAGTGTTCATGAGCAGCTGCTCATGACCTCAAG	1601	
QY	1845	TACAAAAAATCTTTTGTGTAAGTATTTGTCAAGTATTATAAACAACATGCAAGAAATAT	1904	
DB	1602	TATAAAGAGCTGTTCGGCTTCGATTTGCTTAAAACTATGAGCGTTTTCGACAGAGATTAT	1661	

FEATURES
source

QY	1905	ATCAGTGAATGATCATGACAGAAAGTATCTCTATAACTGCACATTTTCAGTTTCAGATGTTTACT	1961
Db	1662	GTGACAGATGACCCAGACAGAGAGATTTTTCAGTCGCAGACCTCTCGGTTTCAGATATTCACA	1721
QY	1965	GTTCTACTCTGGCTCGACATCTTTATGAAGACGAGAAATGTTATCTCTGTCTATTCAGTAA	2024
Db	1722	GTTCTCTCACTTGGCGGAATGCTCCTCACAGAGAGAAACCTGATGACCGTATATCATTAAG	1781
QY	2025	ACTCTGCTAGAAAGTTTACCT- ----GAGTACTTGGACAGGAACAATAAATCAACTTC	2078
Db	1782	GCTTTCATGGACCAATTTGAACACAGAGATGCCCAGGCGAGATCCAGTTTGAACGCTAC	1841
QY	2079	CAGGGTTATAGCCAGGACAAATTTGGGAAGAGTATATGACGTAATATGTGACCTAAAGTAT	2138
Db	1842	ACTGCCCTCCAAAGCCTTCAAGTTTCAGGAGAGTCAGAGCCTCATCTTAGATCTCAAGTAT	1901
QY	2139	ATCCTGATCAGCAACCCACAATATGAGCAAGAAAGATTAAAGATGCAATTCCTTGAAGGT	2198
Db	1902	GTATTGATTAGCAAAACCAAGGAGTGTGCAGATGAGCTGAGGCGAAGTCTTTACRAAGG	1961
QY	2199	TTTTCGATCTTTTTCAGATCTTTACCTGTATGCGAGGGAATGGAANAATCCGAAACAG	2258
Db	1962	TTGCATGCCCTTTTGGAAATTTACTTGAAGTGATGCGAGGGAATGGACCCGATCAACCGCTCAG	2021
QY	2259	GTTGGGCAACATTTGAAGTGGATCCTGATTTGGGAGGCTGCCATTGCTATACAGATGCAA	2318
Db	2022	GTGGGACAGCATTTGAGATGGAGCCAGAGTGGGAAGCAGCTTCACACTGCGAGATGAAG	2081
QY	2319	TTGGAAGAATATTTTACTCATGTTCCAAAGAGTGTGTGCTTGATGAAGAAACCTTTACTT	2378
Db	2082	CTGACACAGCTCATCTCAATGGTGCAGGACTGGTGTGCTCGGACGAAAAGTGTTAAT	2141
QY	2379	GTGGCTTATAAAGAATCTCAAAAGCTGTGATGAGGTGCAGTACCAAGTTTTCATATCTAGT	2438
Db	2142	GAAGCTTACAAGAAATGCCGTGGCTGTGCTGCACACAGTGTGATGGCGGATTTACTGATGGT	2201
QY	2439	AGCAAGACAGT- --AGTACAATCGTGTGGACATAGTTTGGAAACAAGTCTCTACAGATGA	2495
Db	2202	GAACAGCAATACACACTCAGTATTTGTGGACACTCGGTGGAAACCATCAGATACTGTGTT	2261
QY	2496	TCTGAGATCTTGTAAACATACATCGCCACTCTTAGGACCCCTGCTGGTCTTCATGTA	2555
Db	2262	TCOCAAGAAAAGTTAGCATTCACCTCCCAATTTCTCGCTTGTGTCAGGTTTGCATGTA	2321
QY	2556	CGTTTAAAGCAGGCTGGGTGCTGTTTCAAGACTGCATGAATTTGTGTCTTTTGGAGGACTT	2615
Db	2322	TTGTTAGCAAAAGTGAAGTGGCATATAAATTTCCAGAGCTCTACTCTTAAGTGAACGTG	2381
QY	2616	CAAGTAGAGTACTAGTGGAAATATCCCTTTACGTTGTCTGTGTTGGTGGCCAGGTTGTT	2675
Db	2382	AGCCCCCATGTTGATAGAACATCTCTTAGATGCTTGTCTTATGTGCTCAAGTGCAT	2441
QY	2676	GCTGAGATGGCGGAAGAAATGGACTGTCTCTTATTAGCCAGGTGTTTATTACCAAGAT	2735
Db	2442	GCTGGGATGGAGAGAAATGCTTCTCTCTAGTAANTCAGATCTATTACTACCATAT	2501
QY	2736	GTTAAGTGCAGAGAANAATGTATGATAAGATATCATCATGCTTCAGATTTGGTGCATCT	2795
Db	2502	GTGAAATGCAGGCGAGAGATGTTCCGACAAGGACATAGTGATGCTTCAGACAGGTGCTCC	2561
QY	2796	TTAATGGATCCCAATAGTTCTTGTGTACTGGTACTTTCAGAGGTATGAACCTTCCCGAGGT	2855
Db	2562	ATGATGACCCAAACCACTTCTGTATGATCATGCTCAGCGGCTTGTGAACCTCTATCAGCTC	2621
QY	2856	TTTAAACAAGACCATATCTACAAA- ----AGACAGGATTTTGATT	2894
Db	2622	TTCAGACGCGCTGACTATGGGAAGAGATTCAGTTCTGAGGTTACCCATAAGGACGCTGTT	2681
QY	2895	AAACAATATATACACTTAATAGAAAGAAATGCTTCAGGTCTCATCTATATTGTGGGTGAG	2954
Db	2682	CACGACACAACTCTGATCGAAGAGATGCTCTACTCATCATCATGCTTTGTGGGAA	2741
QY	2955	CGTTATGTACTGGAGTGGAAATGTGACCAAGAAGAGGCTCACATGAGAGAAATCATTT	3014

RESULT	8
LOCUS	A1929033
DEFINITION	A1929033 682 bp mRNA linear EST 23-AUG-1999 au64c10.yl Schneider fetal brain 00004 Homo sapiens cDNA clone IMAGE:2519538 5' similar to TR:O70481 O70481 UBIOUITIN-PROTEIN LIGASE E3 COMPONENT N-RECOGNIN ; mRNA sequence.
ACCESSION	A1929033
VERSION	A1929033.1 GI:5664997
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

```
REFERENCE 1 (bases 1 to 682)
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Maizra,M., Martin
,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
White,Y., Wylie,T., Waterston,R. and Wilson,R.
TITLE WashU-NCI human EST Project
JOURNAL Unpublished
COMMENT Other_ESTs: au64c10.xl
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Possible reversed clone: similarity on wrong strand
Seq primer: -40RP from Gibco
High quality sequence stop: 450.
FEATURES
    source
        1..682
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:2519538"
            /sex="male"
            /tissue_type="frontal lobe"
            /dev_stage="5 months post-conception"
            /lab_host="DH10B"
            /clone_lib="Schneider fetal brain 00004"
            /note="Organ: brain; Vector: pBluescript SK (Stratagene);
            Site.1: SstI; Site.2: XhoI; Double-stranded cDNA was
            prepared from human fetal brain tissue. 5' and 3'
            adaptors were used in cloning as follows: 5' adaptor
            sequence:
            5'-GAGAGAGAGAGAGCTCAAGGATCTTAATTAATAATCCCCCCCCCC-3'
            and 3' adaptor sequence:
            5'-CAGAGAGAGACTCGAGTTTCTTTTCTTTT-3'. The library was
            size-selected for >0.5 kb inserts and has an average
            insert size estimated at 1.2 kb. This library was
            constructed using the CAP-trapper method for full-length
            enrichment and has not undergone amplification. Library
            was constructed by Dr. Claudio Schneider (LNCIB-Area
            Science Park, Trieste, Italy)."
BASE COUNT 178 a 155 c 166 g 182 t 1 others
ORIGIN
Query Match 10.7%; Score 674.6; DB 9; Length 682;
Best Local Similarity 99.3%; Pred. NO. 1.8e-128;
Matches 677; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy 5261 AACTCCGCGCTGAGGAAGTCAATTCGAGAGAGAGTACAGTGCACCTGTAG 5320
Db 1 AACTCCGCGCTGAGGAAGTCAATTCGAGAGAGAGTACAGTGCACCTGTAG 60
Qy 5371 CTATCTATCTTACCTACAAATTTGCTCTCTCCAGGAATATTCGGATACGTAA 5380
Db 61 CTATCTATCTTACCTACAAATTTGCTCTCTCCAGGAATATTCGGATACGTAA 120
Qy 5381 GCCCTTGTCTCAGAGGTGGTGTGAGATCCCTGCGCTTACTAACTGTTTGAAGCAAAAAA 5440
Db 121 GCCCTTGTCTCAGAGGTGGTGTGAGATCCCTGCGCTTACTAACTGTTTGAAGCAAAAAA 180
Qy 5441 CACCGTGGTCAAGTACCTAGAAAAAGAAATAGTTTGATAGAGCTTCTGTAGTACTATAG 5500
Db 181 CACCGTGGTCAAGTACCTAGAAAAAGAAATAGTTTGATAGAGCTTCTGTAGTACTATAG 240
Qy 5501 CTGGCTCTGATCAAGCTTCTCATTTTCAGGTGCCACGGTCTGCAGATGATGACGGAAA 5560
Db 241 CTGGCTCTGATCAAGCTTCTCATTTTCAGGTGCCACGGTCTGCAGATGATGACGGAAA 300
Qy 5561 GCATCTCTCTCTGCTCTTTCTGTGGGCTATACTATGTTCTCAGAACATTTGCTGCCA 5620
Db 1 GCATCTCTCTCTGCTCTTTCTGTGGGCTATACTATGTTCTCAGAACATTTGCTGCCA 11
301 GCATCTCTCTCTGCTCTTTCTGTGGGCTATACTATGTTCTCAGAACATTTGCTGCCA 360
5621 GGAAATTTGGAACGGGGAAGAGTTGGAGCTTGCATTTTTCAGCAGCTTTCACCTGTGAGC 5680
361 GGAAATTTGGAACGGGGAAGAGTTGGAGCTTGCATTTTTCAGCAGCTTTCACCTGTGAGC 420
5681 CGGAGTCTGCATTTTCCCTAAAAATCAGAAATGCCGAGTGGTCTCTGGTTGAAGGTAAGC 5740
421 CGGAGTCTGCATTTTCCCTAAAAATCAGAAATGCCGAGTGGTCTCTGGTTGAAGGTAAGC 480
5741 CAGAGGCTGTGCTATCCAGCTCCCTTACTTGGATGAATATGGAGAACACAGCCCTGGCCT 5800
481 CAGAGGCTGTGCTATCCAGCTCCCTTACTTGGATGAATATGGAGAACACAGCCCTGGCCT 540
5801 GAAGAGGGCAACCCCTTTCATTTATCTCTGAGCGGTATCGGAAGCTCCATTTGGTCTG 5860
541 GAAGAGGGCAACCCCTTTCATTTATCTCTGAGCGGTATCGGAAGCTCCATTTGGTCTG 600
5861 GCAACAACACTGCATTATAGAGAGATTGCTAGGAGCCAGACGACTAATCAGATGTTATT 5920
601 GCAACAACACTGCATTATAGAGAGATTGCTAGGAGCCAGACGACTAATCAGATGTTATT 660
5921 TGGATTCACTGCGAGTTACTG 5942
661 GGGATTCACTGCGAGTTACTG 682
RESULT 9
LOCUS BUI55953 898 bp mRNA linear EST 03-SEP-2002
DEFINITION AGENCOURT_7966363 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6165232
5', mRNA sequence.
ACCESSION BUI55953
VERSION BUI55953.1 GI:22669485
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 898)
NIH-MGC http://mgi.nci.nih.gov/.
NATIONAL INSTITUTES OF HEALTH, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-r@mail.nih.gov
Tissue Procurement: ATCC/DCTP/DTP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13523 row: j column: 17
High quality sequence stop: 606.
FEATURES
    Location/Qualifiers
        1..898
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:6165232"
            /tissue_type="melanotic melanoma"
            /lab_host="DH10B (phage-resistant)"
            /clone_lib="NIH_MGC_72"
            /notes="Organ: skin; Vector: pCMV-SPORT6; Site.1: NotI;
            Site.2: SalI; Cloned unidirectionally. Primer: Oligo dT.
            Average insert size 2 kb. Library constructed by Life
            Technologies."
BASE COUNT 254 a 181 c 191 g 272 t
ORIGIN
Query Match 10.6%; Score 668.8; DB 13; Length 898;
Best Local Similarity 99.6%; Pred. No. 3e-127;
Matches 681; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
```

QY	5626	TTGTGAACGGGGAAGAGTTGGAGCTTGGACATTTTTCACGCATCTTCACTGTGGAGCCGGAG	5685
Db	12	TTGTGAACGGGGAAGAGTTGGAGCTTGGACATTTTTCACGCATCTTCACTGTGGAGCCGGAG	71
QY	5686	TCGTGCATTTTCCTAAAAATCAGAGAAATGCCGAGTGGTCCCTGGTTGAAGGTAAGCCAGAG	5745
Db	72	TCGTGCATTTTCCTAAAAATCAGAGAAATGCCGAGTGGTCCCTGGTTGAAGGTAAGCCAGAG	131
QY	5746	GCTGTGCCTATCCAGCTCCTTACTTGGATGAATATGAGAAACAGACCCCTGGCTGAAGA	5805
Db	132	GCTGTGCCTATCCAGCTCCTTACTTGGATGAATATGAGAAACAGACCCCTGGCTGAAGA	191
QY	5806	GGGGCAACCCCTTCATTTATCTCGTGCAGCGGTATCGGAAGCTCCATTTGGTCTGGCAAC	5865
Db	192	GGGGCAACCCCTTCATTTATCTCGTGCAGCGGTATCGGAAGCTCCATTTGGTCTGGCAAC	251
QY	5866	AACACTGCATTTATAGAGAGATTCTAGGAGCAAGAGACTAATCAGATGTTATTGGAT	5925
Db	252	AACACTGCATTTATAGAGAGATTCTAGGAGCAAGAGACTAATCAGATGTTATTGGAT	311
QY	5926	TCAACTGGCAGTTACTGTGAGCTCCAACTCTGCCTCAAGACAATCACAATGACGACAGT	5985
Db	312	TCAACTGGCAGTTACTGTGAGCTCCAACTCTGCCTCAAGACAATCACAATGACGACAGT	371
QY	5986	AGTAAAGCGTGATTCAAAATTATGGAACCTTTCTGAGGCGCTGGGAAGTATTGGAGGGT	6045
Db	372	AGTAAAGCGTGATTCAAAATTATGGAACCTTTCTGAGGCGCTGGGAAGTATTGGAGGGT	431
QY	6046	CTTTTCTCCATGCTCCAGGTTCACTTACATCAATAAATATTCTTAATGAGTATTGCT	6105
Db	432	CTTTTCTCCATGCTCCAGGTTCACTTACATCAATAAATATTCTTAATGAGTATTGCT	491
QY	6106	TTCAATTAGCAACATATGCTTCACAGGAAAAAGGACATAGATCAATCTGTTTATGTG	6165
Db	492	TTCAATTAGCAACATATGCTTCACAGGAAAAAGGACATAGATCAATCTGTTTATGTG	551
QY	6166	CTAGTATTTCCAGGAATTTATCCCTTCATAATTTGCTCATTTTCAATTTTATTCATCC	6225
Db	552	CTAGTATTTCCAGGAATTTATCCCTTCATAATTTGCTCATTTTCAATTTTATTCATCC	611
QY	6226	ACTTGTGATGTAAGTCAAGCTCAACAGTTGTAAGATTTTATGCTTTTGGTAACTCTT	6284
Db	612	ACTTGTGATGTAAGTCAAGCTCAACAGTTGTAAGATTTTATGCTTTTGGTAACTCTT	671
QY	6285	CTGCAATTTTGTATTTGGTGTCTT	6308
Db	672	CTGCCATTTGTATTTGGGGTTT	695

RESULT 10
BU618516/c
LOCUS
DEFINITION
UI-H-FHI-bfk-c-21-0-UI-s1 NCI_CGAP_FHI Homo sapiens CDNA clone
UI-H-FHI-bfk-c-21-0-UI 3', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 668)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: James Martin
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Clone distribution information can be obtained from Dr. M. Bento Soares, bento-soares@uiowa.edu
Seq primer: M13 FORWARD
POLYA=Yes.

FEATURES
source
Location/Qualifiers
1. 668
/organism="Homo sapiens"
/mol_type="mrna"
/db_xref="taxon:9606"
/clone="UI-H-FHI-bfk-c-21-0-UI"
/tissue_type="Cell Line"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI_CGAP_FHI"
/note="Organ: Chondrosarcoma; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site:1: EcoR I; Site:2: Not I; NCI_CGAP_FHI is a normalized cDNA library obtained from a cell line derived from grade I chondrosarcoma tissue. The library was constructed and normalized according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is AGAATCCGGC. The cell line was provided by Dr. James Martin from the University of Iowa."
TAG_L1B=UI-H-FHI
TAG_TISSUE=Human Chondrosarcoma Cell Line CS8 - Grade 1
Chondrosarcoma
TAG_SEQ=AGAATCCGGC"
BASE COUNT 180 a 143 c 110 g 235 t
ORIGIN

Query Match 10.5%; Score 664.8; DB 13; Length 668;
Best Local Similarity 99.7%; Pred. No. 1.9e-126;
Matches 666; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY	2607	GAGGACTTTCAGTAGAGGTACTAGTGGAAATATCTTTACGTGCTGTTGGTGGTGGC	2666
Db	668	GAGGACTTTCAGTAGAGGTACTAGTGGAAATATCTTTACGTGCTGTTGGTGGTGGC	609
QY	2667	CAGGTTGCTGCTGAGATGTGGCAAGAAATGACTGCTCTATTAGCCAGGTTGTTTAT	2726
Db	608	CAGGTTGCTGCTGAGATGTGGCAAGAAATGACTGCTCTATTAGCCAGGTTGTTTAT	549
QY	2727	TACCAAGATGTTAAGTGCAGAGAGAAATGTATGATAAGATATCATGCTTCAGATT	2786
Db	548	TACCAAGATGTTAAGTGCAGAGAGAAATGTATGATAAGATATCATGCTTCAGATT	489
QY	2787	GGTGATCTTTAATGGATCCCAATAGTTCTTGTACTGCTACTTCAGAGGTATGAAT	2846
Db	488	GGTGATCTTTAATGGATCCCAATAGTTCTTGTACTGCTACTTCAGAGGTATGAAT	429
QY	2847	GCCGAGGCTTTTAAACAGACCATATCTACAAAGACCAGGATTGATTAAACAATAAT	2906
Db	428	GCCGAGGCTTTTAAACAGACCATATCTACAAAGACCAGGATTGATTAAACAATAAT	369
QY	2907	ACACTAATAGAGAAATGCTTCAGGCTCCTCATCTATTGTTGGGTGAGGCTTATGACCT	2966
Db	368	ACACTAATAGAGAAATGCTTCAGGCTCCTCATCTATTGTTGGGTGAGGCTTATGACCT	309
QY	2967	GGAGTGGGAAATGTGACCAAGAGAGGTGTCACAAATGAGAGAAATCATTCACCTTGC	3026
Db	308	GGAGTGGGAAATGTGACCAAGAGAGGTGTCACAAATGAGAGAAATCATTCACCTTGC	249
QY	3027	ATTGAACCCATGCCACACAGTGCCTATTCGCAAAATTTACCTGAGAAATGAAATATGAA	3086
Db	248	ATTGAACCCATGCCACACAGTGCCTATTCGCAAAATTTACCTGAGAAATGAAATATGAA	189

```

QY 3087 ACTGCTTAGAGATGTCATAAACAAGTGGCCACATTTAAGAAACCGAGGTGTATCAGGC 3146
Db 188 ACTGCTTAGAGATGTCATAAACAAGTGGCCACATTTAAGAAACCGAGGTGTATCAGGC 129
QY 3147 CATGAGTATTGAACATAAAGATGAATCACTGAAAGACTTCAATATGTACTTTTATCAT 3206
Db 128 CATGAGTATTGAACATAAAGATGAATCACTGAAAGACTTCAATATGTACTTTTATCAT 69
QY 3207 TACTCCAAACCCAGCATAGCAAGGCTGAACATATGCAGAGAGAGAGAGAGAGAGAGAG 3266
Db 68 TACTCCAAACCCAGCATAGCAAGGCTGAACATATGCAGAGAGAGAGAGAGAGAGAGAG 9
QY 3267 AACAAAGA 3274
Db 8 AACAAAA 1

RESULT 11
BG534574 756 bp mRNA linear EST 03-APR-2001
LOCUS 602553425F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4663182 5',
DEFINITION mRNA sequence.
ACCESSION BG534574
VERSION BG534574.1 GI:13526116
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 756)
NIH-MGC http://mgs.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabs@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLC1465 row: a column: 07
High quality sequence stop: 751.
FEATURES
Location/Qualifiers
1..756
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4663182"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_77"
/site="organ: lung; Vector: pDNR-LIB (clontech); Site:1:
SfiI (ggccctcgcc); Site:2: SfiI (ggccattggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor sequence:
5'-ATTCAGAGCGGCGGCGGCACATG-dT(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.9
kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH_MGC Library."
BASE COUNT 274 a 138 c 152 g 192 t
ORIGIN

Query Match
Best Local Similarity 10.4%; Score 656.6; DB 10; Length 756;
Matches 730; Conservativity 0; Mismatches 19; Indels 12; Gaps 5;

QY 2836 GGTATGAATTCGCCGAGCTTTTAAACAAGACCATATCTACAAAGACCAAGGAGGATTTGATTA 2895
Db 1 GGTATGAATTCGCCGAGCTTTTAAACAAGACCATATCTACAAAGACCAAGGAGGATTTGATTA 60

```

```

QY 2896 AACATATATACACTAATAGAGAATAATGCTTCAGGTCCCTCATCTATATTTGTTGGTGAGC 2955
Db 61 AACATATATATACACTAATAGAGAATAATGCTTCAGGTCCCTCATCTATATTTGTTGGTGAGC 120
QY 2956 GTTATCTACTGAGTGGGAAATGTGACCAAAAGAGGTGACAAATCAGAGAAATCATTC 3015
Db 121 GTTATCTACTGAGTGGGAAATGTGACCAAAAGAGGTGACAAATCAGAGAAATCATTC 180
QY 3016 ACTTGTCTTGCATTGAACCCATGCCACACAGTGCATTGGCAAAAATTTACCTGAGAAATG 3075
Db 181 ACTTGTCTTGCATTGAACCCATGCCACACAGTGCATTGGCAAAAATTTACCTGAGAAATG 240
QY 3076 AAAATATGAACCTGGCTTAGAGAAATGTCTATAACAAGTGGCCACATTTAAGAAACCAAG 3135
Db 241 AAAATATGAACCTGGCTTAGAGAAATGTCTATAACAAGTGGCCACATTTAAGAAACCAAG 300
QY 3136 GTGTATCAGGCCCATGGAGTTTATGAACATAAAGATGAATCACTGAAAGACTTCAATATGT 3195
Db 301 GTGTATCAGGCCCATGGAGTTTATGAACATAAAGATGAATCACTGAAAGACTTCAATATGT 360
QY 3196 ACTTTTATCATTTACTCCAAAACCCAGCATAGCAAGGCTGAACATATGCAGAGAGAGAGAG 3255
Db 361 ACTTTTATCATTTACTCCAAAACCCAGCATAGCAAGGCTGAACATATGCAGAGAGAGAGAG 420
QY 3256 GAAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3309
Db 421 GAAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
QY 3310 GCCCTGCTTTCAGCAAAAGTGAATTAACCTTCTCAACTGTGATAT-CATGATGTACATTTCTC 3368
Db 481 GCCCTGCTTTCAGCAAAAGTGAATTAACCTTCTCAACTGTGATATCCATGATGTACATTTCTC 540
QY 3369 AGGACCGTATTTCAGCGGGCAATAGACACAGATTCTAACTTGTGGACCGAAGGATGCTC 3428
Db 541 AGGACCGTATTTCAGCGGGCAATAGACACAGATTCTAACTTGTGGACCGAAGGATGCTC 600
QY 3429 CAAATGGCTTTTCATATCTGCGATTGGGTTTACTAGAAGAGAGCAACAGCTTCAAAAA 3488
Db 601 CAAATGGCTTTTCATATTTCTGCGATTGGGTTTACTAGAAGAGAGCAACAG-TTCAAAAA 659
QY 3489 GCTCTCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3548
Db 660 GCTCTCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 719
QY 3549 GCCATGAATATACAAATGCTTTTGGAAAACTCAAGGAAT 3589
Db 720 G-CATGAATATACAA---TGTTTTGGAAAAATCAAGGATT 756

RESULT 12
AI361043/c 641 bp mRNA linear EST 15-FEB-1999
LOCUS qy03fil.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:2010957 3',
DEFINITION similar to TR:P91133 P91133 SIMILAR TOS. CEREBISIAE
UBIQUITIN-PROTEIN LIGASE E3 COMPONENT Sp:P19812. ;, mRNA sequence.
ACCESSION AI361043
VERSION AI361043.1 GI:4112664
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 641)
NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/BTGP), Tumor Gene Index
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabs@mail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima

```

LOCUS	DEFINITION	AGENCOURT_10609014	NIH_MGC_126	Homo sapiens	cdna clone	linear	EST
LOCUS	DEFINITION	AGENCOURT_10609014	NIH_MGC_126	Homo sapiens	cdna clone	linear	EST
		IMAGE:6727400	5'	mrna	sequence		
		BU55331	7/4 bp	mrna			
		BU55331		mrna			
		AGENCOURT_10609014	NIH_MGC_126	Homo sapiens	cdna clone	linear	EST
		IMAGE:6727400	5'	mrna	sequence		
		BU55331	7/4 bp	mrna			
		BU55331		mrna			

DEFINITION
IMAGE:6727400 5', mRNA sequence.
ACCESSION BU955331
AGNC0000010000909014 NIN_MOC_120 homo sapiens cdna clone

ACCESSION	BU955331
VERSION	BU955331.1
	GI:24184903

```
VERSION BU955331.1 GI:24184903  
KEYWORDS EST.  
*****
```

KEYWORDS EST. Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens

ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS
1 (bases 1 to 774)
NIH-MGC <http://mgc.ncbi.nih.gov/>.
National Institutes of Health, Bethesda, Maryland, USA

JOURNAL
 TITLE
 NATIONAL INSTITUTES OF HEALTH, MAMMALIAN GENE COLLECTION (MGC)
 UNPUBLISHED
 COMMENT
 CONTACT: ROBERT STRAUSBERG, Ph.D.

Email: cgapbs-r@mail.nih.gov
Tissue Procurement: NCI

cdNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
cdNA Library Preparation: Michael Brownstein Laboratory
DNA Sequencing by: Agencourt Bioscience Corporation

Cloning and sequencing by Agencourt Bioscience Corporation. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:

http://image.llnl.gov
Plate: LLCM3047 row: f column: 07
nick quality: accurate step: 555

```

FEATURES
  Location/Qualifiers
    source
      1. .774
    high quality sequence stop: 333.

```

```
/organism="Homo sapiens"  
/mol_type="mRNA"  
/size="406"
```

```

/clone="IMAGE:6727400"
/tissue_type="mixed (pool of 40 RNAs)"
/seq_xref="taxon:9606"

```

```

/clone.lib="NIH_MGC_126"
/lab_host="DH10B (T1-phage-resistant)"

```

```

/note="Vector: pDNR-LIB; Site_1: Sfil (ggccattatggcc);
Site_2: Sfil (ggcgcctcggcc); Double-stranded cDNA was
prepared from a pool of 40 cell line polyA+ RNAs (bladder

```

2%, blood - 33.4%, brain - 5.6%, breast - 12.5%, colon - 4%, connective tissue - 1.4%, eye - 1%, intestine - 2.6%,

kidney - 2.2%, liver - 5.7%, lung - 10.8%, NK-cell - 5.2%,
ovary - 4%, pharynx - 2.5%, prostate - 4.3%, salivary
gland - 1.3% and skin - 2.3%. 5' and 3' adaptors were

grand 1.5% and skin 2.5%). 3' and 5' adaptors were used in cloning as follows:
5'-AAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCGGG-3' and

5'-ATTCTAGGGCGAGGGCCGACATG-dT(30)NN-3'. Full-length enriched library was constructed using the Clontech

kb size fraction (other fractions present in NIH_MGC_127 and NIH_MGC_128). Library created in the laboratory of T. Creator SMART kit and size-selected to contain the 0.5-1

Usdin, M.D., Ph.D. (NIMH, NIH). Note: this is a NIH_MGC Library."

BASE COUNT	247 a	151 c	154 g	222 t
ORIGIN				

Query Match	10.1%	Score 635.2;	DB 13;	Length 774;
Best Local Similarity	95.2%	Pred. No. 2.5e-120;		

Matches 678; Conservative : 0; Mismatches 28; Indels 6; Gaps 2;
O::
1375 AATTCACGCGCCGGCAGTAAATGCACCAAAACACCACTCATATACACCAATACAAA 1434

QY 1375 ATTGTGTCCTTTTCATATGATGAACACCATTCATATGACCCCTCATATACAGCTACAAA 62

QY 1435 GAGCTCTTGACTGTGAGCTCGACGAGGCCAGTTGCATACCACTGCCATTGACAAAGAGG 1494

Db 63 GAGCTCTTGACTGTGAGCTCGCAGAGAGCCCGAGTTGCATACCACTGCCATTGCACAAAGAGG 122

QY	1495	GTGTCGGGCTGTTAAAGCGGGAGCTTATGCTTGCCAGGAAGCAAAAGAGATATAA	1554
Db	123	GTGTCGGGCTGTTAAAGCGGGAGCTTATGCTTGCCAGGAAGCAAAAGAGATATAA	182
QY	1555	AGAGTCATTACAGAAAATGCTCTCAACATCCACTTCTAGTAGAGTATTACACTCAGAGA	1614
Db	183	AGAGTCATTACAGAAAATGCTCTCAACATCCACTTCTAGTAGAGTATTACACTCAGAGA	242
QY	1615	TATGGCTCATCAGAAAATTTGCCTTGGCTTGGTTCCTCGATGCAACAAAATATGAGCT	1674
Db	243	TATGGCTCATCAGAAAATTTGCCTTGGCTTGGTTCCTCGATGCAACAAAATATGAGCT	302
QY	1675	ATTCAAGTGACTTTTAGGCAGATCTTTTGGCCAAGCATGCCCTTAGAGAAGAACCTGACTCGG	1734
Db	303	ATTCAAGTGACTTTTAGGCAGATCTTTTGGCCAAGCATGCCCTTAGAGAAGAACCTGACTCGG	362
QY	1735	AGAAATCCCTCTCATTAAGCAGGTTAATGCTTTGGGATGCAAGACTTTATAAGGTGCC	1794
Db	363	AGAAATCCCTCTCATTAAGCAGGTTAATGCTTTGGGATGCAAGACTTTATAAGGTGCC	422
QY	1795	GTAAGATCCCTTCATGAATTCATCTTCAGCAGTTTTTTTATGGAGATGGAATACAAAAAC	1854
Db	423	GTAAGATCCCTTCATGAATTCATCTTCAGCAGTTTTTTTATGGAGATGGAATACAAAAAC	482
QY	1855	TCCTTGCTATGGAATTTGTCAGATATTATAACAACATCGCAAGAAGATATATCAGTGATG	1914
Db	483	TCCTTGCTATGGAATTTGTCAGATATTATAACAACATCGCAAGAAGATATATCAGTGATG	542
QY	1915	ATCATGACAGAAGTATCTCTATACCTGCATCTTCAGTTCAGATGTTTACTGTTCTCTACT	1974
Db	543	ATCATGACAGAAGTATCTCTATACCTGCATCTTCAGTTCAGATGTTTACTGTTCTCTACT	602
QY	1975	TGGCTCCACATCTTATTGAAGACGAGAAATGTTATCTCTGTCATTACTGAACTCTGCTAG	2034
Db	603	TGGCTCCACATCTTATTGAAGACGAGAAATGTTATCTCTGTCATTACTGAACTCTGCTAG	659
QY	2035	AA---GTTTTCACCTGACTCTTGACAGGAAACATAATTAATCACTCCAGGG	2083
Db	660	AAAGTTTTTACCCTGAGTACTTGACCGGGAACAATACAAAATTTTCACATGG	711
RESULT 14			
BG862813			
LOCUS	60279907Af1 NCI_CGAP_Mam4 Mus musculus cdna clone IMAGE:4934370 5'	797 bp mRNA linear EST 29-MAY-2001	
DEFINITION	mRNA sequence.		
ACCESSION	BG862813	GI:14213351	
VERSION	BG862813.1	EST.	
KEYWORDS	Mus musculus (house mouse)		
SOURCE	Mus musculus		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	NH-Bases 1 to 797		
AUTHORS	National Institutes of Health, Mammalian Gene Collection (MGC)		
TITLE	Unpublished		
JOURNAL	Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth Ph.D.		
COMMENT	CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM10863 row: d column: 19 High quality sequence stop: 690. Location/Qualifiers 1..797 /organism="Mus musculus"		
FEATURES			
source			

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2003, 19:40:47 : Search time 12240.1 Seconds
(without alignments)
17396.489 Million cell updates/sec

Title: US-09-724-126A-18
Perfect score: 5205
Sequence: 1 atggcgacgagagctgg.....tcaactggcagttactgtga 5205

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba:**
- 2: gb_hgt:**
- 3: gb_in:**
- 4: gb_ov:**
- 5: gb_ov:**
- 6: gb_pat:**
- 7: gb_ph:**
- 8: gb_pl:**
- 9: gb_pr:**
- 10: gb_ro:**
- 11: gb_sts:**
- 12: gb_sy:**
- 13: gb_un:**
- 14: gb_vi:**
- 15: em_ba:**
- 16: em_fun:**
- 17: em_hum:**
- 18: em_in:**
- 19: em_mu:**
- 20: em_om:**
- 21: em_or:**
- 22: em_ov:**
- 23: em_pat:**
- 24: em_ph:**
- 25: em_pl:**
- 26: em_ro:**
- 27: em_sts:**
- 28: em_un:**
- 29: em_vi:**
- 30: em_hgt_hum:**
- 31: em_hgt_inv:**
- 32: em_hgt_other:**
- 33: em_hgt_mus:**
- 34: em_hgt_pln:**
- 35: em_hgt_rtd:**
- 36: em_hgt_mam:**
- 37: em_hgt_vrt:**
- 38: em_sv:**
- 39: em_hgt_hum:**
- 40: em_hgt_mus:**
- 41: em_hgtg_other:**

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	5136	98.7	5250	9	AY061886	AY061886 Homo sapi
2	4971.6	95.5	5128	9	AF525401	AF525401 Homo sapi
3	4085.6	78.5	6395	6	AR030784	AR030784 Sequence
4	4085.6	78.5	6395	6	AR121463	AR121463 Sequence
5	4085.6	78.5	6395	10	AF061555	AF061555 Mus muscu
6	2927.4	56.2	3059	6	AX714232	AX714232 Sequence
7	2927.4	56.2	3059	9	AK056441	AK056441 Homo sapi
8	2434.4	46.8	2550	6	BD156870	BD156870 Primer fo
9	2434.4	46.8	2550	9	AK027803	AK027803 Homo sapi
10	1294.2	24.9	5268	9	AY061884	AY061884 Homo sapi
11	1151.8	22.1	5265	10	AY061885	AY061885 Mus muscu
12	996.2	19.1	1001	6	AR030785	AR030785 Sequence
13	996.2	19.1	1001	6	AR121464	AR121464 Sequence
14	994.2	19.1	999	9	AF061556	AF061556 Homo sapi
15	979.4	18.8	6158	9	AB002347	AB002347 Human mRN
16	771	14.8	3502	10	BC031403	BC031403 Mus muscu
17	650	12.5	2958	9	AK026998	AK026998 Homo sapi
18	615.4	11.8	818	9	HSX521	BD149613 Primer fo
19	534.6	10.3	712	6	BD149613	BD149613 Primer fo
20	513.4	9.9	2475	10	BC026391	BC026391 Mus muscu
21	486.2	9.3	2512	9	AK026948	AK026948 Homo sapi
22	313	6.0	1368	10	BC025617	BC025617 Mus muscu
23	258.8	5.0	190727	9	AC068724	AC068724 Homo sapi
24	246.8	4.7	807	6	BD079625	BD079625 Cancer-as
25	224	4.3	5027	10	AK122254	AK122254 Mus muscu
26	215.2	4.1	910	10	MM1UBR2	AF067372 Mus muscu
27	215.2	4.1	168200	2	AC120631	AC120631 Rattus no
28	215.2	4.1	198946	10	AL844548	AL844548 Mouse DNA
29	215.2	4.1	235978	2	AC094209	AC094209 Rattus no
30	212.2	4.1	107304	2	AC016274	AC016274 Homo sapi
31	212.2	4.1	164468	2	AC021899	AC021899 Homo sapi
32	212.2	4.1	166518	9	AC090514	AC090514 Homo sapi
33	195	3.7	6381	3	AY094815	AY094815 Drosophil
34	191.6	3.7	800	6	BD079626	BD079626 Cancer-as
35	180.8	3.5	164468	2	AC021899	AC021899 Homo sapi
36	160.6	3.1	3980	9	HS1UBR4	AF067383 Homo sapi
37	160.2	3.1	3411	9	AK090874	AK090874 Homo sapi
38	159	3.1	455	6	BD071823	BD071823 Secreted
39	157.6	3.0	594	11	G79301	G79301 S210P604SRG
40	150.2	2.9	107304	2	AC016274	AC016274 Homo sapi
41	146.2	2.8	181602	10	AL935168	AL935168 Mouse DNA
42	132	2.5	28834	2	AC017982	AC017982 Drosophil
43	132	2.5	152545	3	AC010922	AC010922 Drosophil
44	132	2.5	179016	3	AC018489	AC018489 Drosophil
45	132	2.5	298641	3	AE003504	AE003504 Drosophil

ALIGNMENTS

RESULT 1
AY061886
LOCUS AY061886 5250 bp mRNA linear PRI 30-DEC-2002
DEFINITION Homo sapiens ubiquitin ligase E3 alpha-I mRNA, complete cds.
ACCESSION AY061886
VERSION AY061886.1 GI:27434483
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 5250)
TITLE Full-length human ubiquitin ligase E3 alpha-I (E3 alpha)
JOURNAL Unpublished

Db	1381	 TTCCAGGGTATAGCCAGGACAAATTCGGAGAGTATATACGATATATATGTGACCTAAAG	1440
Qy	1396	TATATCCTGTATCAGCAAAACCCACAATATGGACAGAAAGATTAAAGATGCAAGTTCCCTTGAA	1455
Db	1441	TATATCCTGTATCAGCAAAACCCACAATATGGACAGAAAGATTAAAGATGCAAGTTCCCTTGAA	1500
Qy	1456	GGTTTTCGATCTTTTTTGAAGATCTTTACCTGTATGCAGGGAATGGAAGAAATCCGAAGA	1515
Db	1501	GGTTTTCGATCTTTTTTGAAGATCTTTACCTGTATGCAGGGAATGGAAGAAATCCGAAGA	1560
Qy	1516	CAGTTTGGCAACACATTTGAAGTGGATCCCTGATTTGGAGGCTGCCATTCGTATACAGATG	1575
Db	1561	CAGTTTGGCAACACATTTGAAGTGGATCCCTGATTTGGAGGCTGCCATTCGTATACAGATG	1620
Qy	1576	CAATTGAAGAATATTTTTACTCATGTTCCCAAGAGTGGTGCCTGTCATGAAGAACCTCTTA	1635
Db	1621	CAATTGAAGAATATTTTTACTCATGTTCCCAAGAGTGGTGCCTGTCATGAAGAACCTCTTA	1680
Qy	1636	CTTTGTGCTTATAAAGAATGTCACAAAGCTGTGATGAGTGCAGTACCAAGTTTCATATCT	1695
Db	1681	CTTTGTGCTTATAAAGAATGTCACAAAGCTGTGATGAGTGCAGTACCAAGTTTCATATCT	1740
Qy	1696	AGTAGCAAGACAGTAGTACAAATCGTGTGGACATAGTTTGGAAACAAAGTCCACAGATA	1755
Db	1741	AGTAGCAAGACAGTAGTACAAATCGTGTGGACATAGTTTGGAAACAAAGTCCACAGATA	1800
Qy	1756	TCGTGAGATCTTGTAAGCATACATCTGCCACTCTCTAGGACCCCTGCTGGTCTTCATGTA	1815
Db	1801	TCGTGAGATCTTGTAAGCATACATCTGCCACTCTCTAGGACCCCTGCTGGTCTTCATGTA	1860
Qy	1816	CGTTTAAAGCAGGCTGGTGCTGTTTCAAGACTGCATGAATTTGTGTCTTTTGAGGACTTT	1875
Db	1861	CGTTTAAAGCAGGCTGGTGCTGTTTCAAGACTGCATGAATTTGTGTCTTTTGAGGACTTT	1920
Qy	1876	CAAGTAGAGTACTAGTGGAAATACCTTTACGTTGTCTGGTGTGGTGCCTGCCAGGTTGT	1935
Db	1921	CAAGTAGAGTACTAGTGGAAATACCTTTACGTTGTCTGGTGTGGTGCCTGCCAGGTTGT	1980
Qy	1936	GCCTCAGATGTCGCGAAGAAATGACATGCTCTCTATTAGCCAGGTGTTTTATACCAAGAT	1995
Db	1981	GCCTCAGATGTCGCGAAGAAATGACATGCTCTCTATTAGCCAGGTGTTTTATACCAAGAT	2040
Qy	1996	GTTTAAAGTCAGAGAAATATGATAAAGATATCATCATGCTTCAGATGCGTGCATCT	2055
Db	2041	GTTTAAAGTCAGAGAAATATGATAAAGATATCATCATGCTTCAGATGCGTGCATCT	2100
Qy	2056	TTAATGGATCCCAATAAGTCTTGTTACTGGTACTTTCAGAGGTATGAACCTGCCAGGCT	2115
Db	2101	TTAATGGATCCCAATAAGTCTTGTTACTGGTACTTTCAGAGGTATGAACCTGCCAGGCT	2160
Qy	2116	TTTTAAACAAGACCATATCTACAAAGACCAGGATTTGATTTAAACAATATAATACATAATA	2175
Db	2161	TTTTAAACAAGACCATATCTACAAAGACCAGGATTTGATTTAAACAATATAATACATAATA	2220
Qy	2176	GAGAAATGCTTCAGGCTCCTCATCTATATTGTGGGTGAGCGTTATGTACTCGAGTGGGA	2235
Db	2221	GAGAAATGCTTCAGGCTCCTCATCTATATTGTGGGTGAGCGTTATGTACTCGAGTGGGA	2280
Qy	2236	AATGTGACCAAGAAGAGTCAATFAGAGAAATCATTCACCTGCTTTGCAATTGAACCC	2295
Db	2281	AATGTGACCAAGAAGAGTCAATFAGAGAAATCATTCACCTGCTTTGCAATTGAACCC	2340
Qy	2296	ATGCCACACAGTGCCATTGCCAAAAATTTACCTGAGATGAAAAATAATGAAGAACTGGCTTA	2355
Db	2341	ATGCCACACAGTGCCATTGCCAAAAATTTACCTGAGATGAAAAATAATGAAGAACTGGCTTA	2400
Qy	2356	GAGAAATGCTATAACAAGGTGCCACATTTAAGAAACCAAGGTGTATCAGGCCATGAGTT	2415
Db	2401	GAGAAATGCTATAACAAGGTGCCACATTTAAGAAACCAAGGTGTATCAGGCCATGAGTT	2460
Qy	2416	TATGAACCTAAAAAGATGAATCACTGAAAGACTTCAATATATGACTTTTATCATTTACTCCAAA	2475

QY 3556 AAATCTCTGCGAATACTGTGATCCCAATATTTCCTTTGCAACCTCAAAAGATAAAGCAGT 3615
|||||
Db 3601 AAATCTCTGCGAATACTGTGATCCCAATATTTCCTTTGCAACCTCAAAAGATAAAGCAGT 3660
|||||
QY 3616 GAGAATGCAGATGCTCTGTCTCAACTTTTGACCCCTGGCAGGGTGATACAGACTGTCTCTG 3675
|||||
Db 3661 GAGATGCAGATGCTCTGTCTCAACTTTTGACCCCTGGCAGGGTGATACAGACTGTCTCTG 3720
|||||
QY 3676 GCCAGAAATACAGTTATATAATAAGACATGCTAAAGAGAGAAACCCAAATTCCTATTTC 3735
|||||
Db 3721 GCCAGAAATACAGTTATATAATAAGACATGCTAAAGAGAGAAACCCAAATTCCTATTTC 3780
|||||
QY 3736 TTTTAATCAAGGAATGGAGATCTTACTTTGGAGTTCCATTCCTCAGTTTGGGGTT 3795
|||||
Db 3781 TTTTAATCAAGGAATGGAGATCTTACTTTGGAGTTCCATTCCTCAGTTTGGGGTT 3840
|||||
QY 3796 GAGTCTTCGATTAAATATCAAAATAGCATCAAGGAAATGGTTATTCCTTTGCCACAACA 3855
|||||
Db 3841 GAGTCTTCGATTAAATATCAAAATAGCATCAAGGAAATGGTTATTCCTTTGCCACAACA 3900
|||||
QY 3856 ATTTATAGAATGGATTGAAGTGCCACCTGATGAAAGGGATCCTCGAGTCCCAATGCTG 3915
|||||
Db 3901 ATTTATAGAATGGATTGAAGTGCCACCTGATGAAAGGGATCCTCGAGTCCCAATGCTG 3960
|||||
QY 3916 ACCTGGAGCACCCTGCTTTCACTATCCAGGCAATGAAATCTATTGGGAGATCAAGGA 3975
|||||
Db 3961 ACCTGGAGCACCCTGCTTTCACTATCCAGGCAATGAAATCTATTGGGAGATCAAGGA 4020
|||||
QY 3976 AAACCTCTGTTGGAGCAGCTTCAAAATAGGAGCAGCATATGCTCTGAAAGCATTAATGCG 4035
|||||
Db 4021 AAACCTCTGTTGGAGCAGCTTCAAAATAGGAGCAGCATATGCTCTGAAAGCATTAATGCG 4080
|||||
QY 4036 TTTTSCAGTTGCACAGAGGATTACCTGCTCCTCAGTCTCTGATACAGAAACATCTGTTCTG 4095
|||||
Db 4081 TTTTSCAGTTGCACAGAGGATTACCTGCTCCTCAGTCTCTGATACAGAAACATCTGTTCTG 4140
|||||
QY 4096 CTTCATATCAGTTGTTCTTCAACATAAAATCAGAAGATACACATGCTCTGCTATATA 4155
|||||
Db 4141 CTTCATATCAGTTGTTCTTCAACATAAAATCAGAAGATACACATGCTCTGCTATATA 4200
|||||
QY 4156 GATCTGTTTCATGTTTGGTGGGTGCTGTGTAGCATTCCTCCATCTCTTATTTGGGATGAC 4215
|||||
Db 4201 GATCTGTTTCATGTTTGGTGGGTGCTGTGTAGCATTCCTCCATCTCTTATTTGGGATGAC 4260
|||||
QY 4216 CCTGTTGATCTGCAGCCTTCTTCAGTTAGTTCTTCTCTATACCAACCTTTATCTCTCCAT 4275
|||||
Db 4261 CCTGTTGATCTGCAGCCTTCTTCAGTTAGTTCTTCTCTATACCAACCTTTATCTCTCCAT 4320
|||||
QY 4276 TTGATCACCATGGCACACATGCTTCAGATACTACTTACAGTAGACAGGCTACCCCTT 4335
|||||
Db 4321 TTGATCACCATGGCACACATGCTTCAGATACTACTTACAGTAGACAGGCTACCCCTT 4380
|||||
QY 4336 GCTCAGGTTCAAGAAGACAGTGAAGAGCTCATTCGCGCATCTCTTTCTTTCGCAAAAT 4395
|||||
Db 4381 GCTCAGGTTCAAGAAGACAGTGAAGAGCTCATTCGCGCATCTCTTTCTTTCGCAAAAT 4440
|||||
QY 4396 TCTCAATATACAAGTGGCTCCATTTGGGTGTGATATTCCTGGCTGGTATTGTTGGGTCTCA 4455
|||||
Db 4441 TCTCAATATACAAGTGGCTCCATTTGGGTGTGATATTCCTGGCTGGTATTGTTGGGTCTCA 4500
|||||
QY 4456 CTGAAGAAATGGCATACCCCTTATCTTCGGCTGTGCTGCATTTGTTTCCACTATTTACTT 4515
|||||
Db 4501 CTGAAGAAATGGCATACCCCTTATCTTCGGCTGTGCTGCATTTGTTTCCACTATTTACTT 4560
|||||
QY 4516 GGGTAACTCCGCTGAGGAAGTGCATACCAATTCGAGAGGAGAGTACAGTCCACTC 4575
|||||
Db 4561 GGGTAACTCCGCTGAGGAAGTGCATACCAATTCGAGAGGAGAGTACAGTCCACTC 4620
|||||
QY 4576 TGTAGCTATCTATCTTACCTACAAATTTGTTCTGCTCTTCCAGGAATATTGGGATACT 4635
|||||
Db 4621 TGTAGCTATCTATCTTACCTACAAATTTGTTCTGCTCTTCCAGGAATATTGGGATACT 4680
|||||

QY 4636 GTAAGGCCCTTGCTCCAGAGCGGTGTGCAGATCCTGCTTACTAACTGTTTGAAGCAA 4695
|||||
Db 4681 GTAAGGCCCTTGCTCCAGAGGTGTGTGCAGATCCTGCTTACTAACTGTTTGAAGCAA 4740
|||||
QY 4696 AAAAACACCGTGGTCCAGTACCCCTAGAAAAAGAAATAGTTTGTATAGAGTTCCTCATGAC 4755
|||||
Db 4741 AAAAACACCGTGGTCCAGTACCCCTAGAAAAAGAAATAGTTTGTATAGAGTTCCTCATGAC 4800
|||||
QY 4756 TATAGTCCCTCCTGTAATCAAGCTTCTCATTTCCAGTGCCACGCTGTGCAGATGAG 4815
|||||
Db 4801 TATAGTCCCTCCTGTAATCAAGCTTCTCATTTCCAGTGCCACGCTGTGCAGATGAG 4860
|||||
QY 4816 CGAAAGCATCCTGCTCTGCTGCTTTCTGTGGGCTATACTATGTTCTCAGAACATTTGC 4875
|||||
Db 4861 CGAAAGCATCCTGCTCTGCTGCTTTCTGTGGGCTATACTATGTTCTCAGAACATTTGC 4920
|||||
QY 4876 TGCCAGGAAATGTGAACGGGGAAGAGTTGGAGCTTGAGATTTGATTTTTCACGCACTTCACTGT 4935
|||||
Db 4921 TGCCAGGAAATGTGAACGGGGAAGAGTTGGAGCTTGAGATTTGATTTTTCACGCACTTCACTGT 4980
|||||
QY 4936 GGAGCCGGAGTCTGCATTTTCTTAAATCAAGAAATGCCGAGTGGTCTGCTTGAAGGT 4995
|||||
Db 4981 GGAGCCGGAGTCTGCATTTTCTTAAATCAAGAAATGCCGAGTGGTCTGCTTGAAGGT 5040
|||||
QY 4996 AAAGCCAGAGGCTGTCCCTATCCAGCTCCTTACTTGGATGAATATGGAGAAACAGACCT 5055
|||||
Db 5041 AAAGCCAGAGGCTGTCCCTATCCAGCTCCTTACTTGGATGAATATGGAGAAACAGACCT 5100
|||||
QY 5056 GGCCTGAAGAGGGGCAACCCCTTCAATTTATCTCGTAGCGGTATCGGAAGCTCCATTTG 5115
|||||
Db 5101 GGCCTGAAGAGGGGCAACCCCTTCAATTTATCTCGTAGCGGTATCGGAAGCTCCATTTG 5160
|||||
QY 5116 GTCCTGGCAACACACTGCATTTATAGAAGATTTCTAGGAGCCAGAGACTATTCAGATG 5175
|||||
Db 5161 GTCCTGGCAACACACTGCATTTATAGAAGATTTCTAGGAGCCAGAGACTATTCAGATG 5220
|||||
QY 5176 TTATTTGGATTCAACTGGCAGTTACTGTGA 5205
|||||
Db 5221 TTATTTGGATTCAACTGGCAGTTACTGTGA 5250
|||||

RESULT 2
AF525401
LOCUS Homo sapiens UBR1 E3a ligase mRNA, partial cds.
DEFINITION AF525401
ACCESSION AF525401
VERSION AF525401.1 GI:27451603
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Cranial; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 5128)
AUTHORS Dgany, O., Avidan, N., Delaunay, J., Krasnov, T., Shalmon, L.,
Shalev, H., Eidelitz-Markus, T., Kapelushnik, J., Cattian, D.,
Pariente, A., Tulliez, M., Cretien, A., Schischmanoff, P.O.,
Iolascon, A., Fibach, E., Koren, A., Rossler, J., Le Merrier, M.,
Yaniv, I., Zaitov, R., Ben-Asher, E., Olender, T., Lancet, D.,
Beckmann, J.S. and Tamary, H.
TITLE Congenital dyserythropoietic anemia type I is caused by mutations
in codanin-1
JOURNAL Am. J. Hum. Genet. 71 (6), 1467-1474 (2002)
MEDLINE 22340442
PUBMED 12434312
REFERENCE 2 (bases 1 to 5128)
AUTHORS Ben-Asher, E.
TITLE Direct Submission
JOURNAL Submitted (27-JUN-2002) Molecular Genetics, The Weizmann Institute
of Science and the Schneider Medical Center, Hertzl Street,
Rehovot 76100, Israel
FEATURES
source l. 5128
/organism="Homo sapiens"

361	TACAAGATGCACTTCTACTGGAGGAGGGTTCTGTGACTGTGTGGAGACACAGAGGCATGG	420
472	AAAACCTGGCCCTTTTGTGTAATATCATGAACCTGGAAGAGCAGGTACTATAAAGAGAAT	531
421	AAAACCTGGCCCTTTTGTGTAATATCATGAACCTGGAAGAGCAGGTACTATAAAGAGAAT	480
532	TCAGCGTGTCCGTTGGAATGAAGAGGTAAATGTCTCAAGCCAGGAAAAATATTTCCCTTCAGTG	591
481	TCAGCGTGTCCGTTGGAATGAAGAGGTAAATGTCTCAAGCCAGGAAAAATATTTCCCTTCAGTG	540
592	ATAAAAATATGTCTGTAGAAAATGACTATATATGGGAAGAGGAAAAAGAACTGCCTCTGAACTC	651
541	ATAAAAATATGTCTGTAGAAAATGACTATATATGGGAAGAGGAAAAAGAACTGCCTCTGAACTC	600
652	CAGATAAGGKNR-----YVC	666
601	CAGTAAGGAGGAAAAATGAAAGATACTATTGTGCTCTTTTCAATGATGAACACCATTC	660
667	VNDHSHDHTCATATACAGCCTACAAGAGCTCTTGACTGTGAGCTCGCAGAGGCCAG	726
661	TATGACCACGTCATATACAGCCTACAAGAGCTCTTGACTGTGAGCTCGCAGAGGCCAG	720
727	TTGCATPACCACTGCCATTGACAAAGAGGGTCTGTGGGCTGTAAAGCGGAGGCTTATGCT	786
721	TTGCATPACCACTGCCATTGACAAAGAGGGTCTGTGGGCTGTAAAGCGGAGGCTTATGCT	780
787	GCATTGCCAGGAACAAAGGAGATATAAGAGATCATTCAGAAAAATGCTCTCAACATCCA	846
781	GCATTGCCAGGAACAAAGGAGATATAAGAGATCATTCAGAAAAATGCTCTCAACATCCA	840
847	CTTCATCTGAAAGTATTACACTCAGACGATTATGGCTCATCAGAAAAATTTGCTTTCGCTCTT	906
841	CTTCATCTGAAAGTATTACACTCAGACGATTATGGCTCATCAGAAAAATTTGCTTTCGCTCTT	900
907	GGTTCTCGGATGAACAAAAATTTAGACTATTCAAGTGACTTTTAGCGAGATCTTTTGCCAA	966
901	GGTTCTCGGATGAACAAAAATTTAGACTATTCAAGTGACTTTTAGCGAGATCTTTTGCCAA	960
967	GCATGCCCTTAGAGNAGAACCTGACTCGGAGAAATCCCTGTCTCATAGCAGGTTAATGCTT	1026
961	GCATGCCCTTAGAGNAGAACCTGACTCGGAGAAATCCCTGTCTCATAGCAGGTTAATGCTT	1020
1027	TGGATGCAAGCTTTATAAGGTGCCGTAAAGATCCCTCATGAATTTGATCTTCAGCAGCT	1086
1021	TGGATGCAAGCTTTATAAGGTGCCGTAAAGATCCCTCATGAATTTGATCTTCAGCAGCT	1080
1087	TTTTTTATGGAGATGGAATACAAAAACTCTTTGCTATGGAATTTGGAAGTATATATAA	1146
1081	TTTTTTATGGAGATGGAATACAAAAACTCTTTGCTATGGAATTTGGAAGTATATATAA	1140
1147	CAACTGCAGAAAGAAATATATCAGTGATGATCATGACAGAGATATCTCTATAACTGCACCTT	1206
1141	CAACTGCAGAAAGAAATATATCAGTGATGATCATGACAGAGATATCTCTATAACTGCACCTT	1200
1207	TCAGTTTCAGATGTTTACTGTTCTCTACTCTCGCTCGACATCTTATTGAAGAGCAGAAATGTT	1266
1201	TCAGTTTCAGATGTTTACTGTTCTCTACTCTCGCTCGACATCTTATTGAAGAGCAGAAATGTT	1260
1267	ATCTCTGCTATTACTGAAACTCTGCTAGAAAGTTTTTACTGAGTACTTTGGACAGGAAACAAT	1326
1261	ATCTCTGCTATTACTGAAACTCTGCTAGAAAGTTTTTACTGAGTACTTTGGACAGGAAACAAT	1320
1327	AAATTCAACTTCAGGGTTATAGCCAGGACAAATTTGGGAAGAGTATATGTCAGTAAATATGT	1386
1321	AAATTCAACTTCAGGGTTATAGCCAGGACAAATTTGGGAAGAGTATATGTCAGTAAATATGT	1380
1387	GACCTAAAGTATATCTCTGATCAGCAAAACCCACAATATGGACAGAAAGATTTAAGAAATCGAG	1446
1381	GACCTAAAGTATATCTCTGATCAGCAAAACCCACAATATGGACAGAAAGATTTAAGAAATCGAG	1440
1447	TTCTTTGAAGGTTTTTCGATCTTTTTTTGAAGATCTTTTACCTGTATGTCAGGGAATGGAAGAA	1506
1441	TTCTTTGAAGGTTTTTCGATCTTTTTTTGAAGATCTTTTACCTGTATGTCAGGGAATGGAAGAA	1500

BASE COUNT	1540 a	1010 c	1137 g	1439 t	2 others
ORIGIN					
Query Match	95.5%	Score 4971.6	DB 9	Length 5128	
Best Local Similarity	98.6%	Pred. No. 0			
Matches 5060	Conservative 11	Mismatches 12	Indels 48	Gaps 4	
QY	52	TTACCCAGACCCCTCAGCGCTCGGCATCTTGTTGGGATCAGCAAGTTGATTTTATACT	111		
Db	1	TTACCCAGACCCCTCAGCGCTCGGCATCTTGTTGGGATCAGCAAGTTGATTTTATACT	60		
QY	112	GCITTTCTTGTCATCTTTGGCACAAATTGGTCCAGAAATTTACTTTGCTGAATGGACCCA	171		
Db	61	GCITTTCTTGTCATCTTTGGCACAAATTGGTCCAGAAATTTACTTTGCTGAATGGACCCA	120		
QY	172	GACTTGAAAGCAGGAGGAAAGTGTACAATGTCAATATTCATCTCACTTGGAAATGGTAC	231		
Db	121	GACTTGAAAGCAGGAGGAAAGTGTACAATGTCAATATTCATCTCACTTGGAAATGGTAC	180		
QY	232	TTATTTGGAGAAGATCCAGATATTTCCTTAGAGAAATTTGAAGCACAGTGGAGCAATTCAG	291		
Db	181	TTATTTGGAGAAGATCCAGATATTTCCTTAGAGAAATTTGAAGCACAGTGGAGCAATTCAG	240		
QY	292	CTTTGTGGGAGGGTTTTCAAAGTGGAGAGACAACCTATTCTTTCGAGGGATTTGTGCAATT	351		
Db	241	CTTTGTGGGAGGGTTTTCAAAGTGGAGAGACAACCTATTCTTTCGAGGGATTTGTGCAATT	300		
QY	352	GATCCAAACATGTGTACTCTGTATGGACTGCTTCCAGGACAGTGTTCATAAAAATCATCGT	411		
Db	301	GATCCAAACATGTGTACTCTGTATGGACTGCTTCCAGGACAGTGTTCATAAAAATCATCGT	360		
QY	412	TACAAGATGCATCTTCTACTTGGAGGAGGGTCTCTGTGACTGTGGAGACACAGAGGCATGG	471		


```
Db 3661 ||||| ACTGTTCTGCCCAATATACAGGTTATATATTAAGACATGCTAAAGAGAAAACCAATT 3720
QY 3727 CCTATTTCTTTAATCAAGGAATGGGAGATTCTACTTTGGAGTTCCATTCCATCCTGAGT 3786
Db 3721 CCTATTTCTTTAATCAAGGAATGGGAGATTCTACTTTGGAGTTCCATTCCATCCTGAGT 3780
QY 3787 TTTGGCGTTGAGTCTCGATTAAATATCAAAATAGCATCAAGAAATGGTTATCTCTTT 3846
Db 3781 TTTGGCGTTGAGTCTCGATTAAATATCAAAATAGCATCAAGAAATGGTTATCTCTTT 3840
QY 3847 GCCACAACAATTTATAGAATTGGATTGAAGTGCACCTGATGAAAGGATCCTCGAGTC 3906
Db 3841 GCCACAACAATTTATAGAATTGGATTGAAGTGCACCTGATGAAAGGATCCTCGAGTC 3900
QY 3907 CCCATGCTGACCTGGAGCACCTCGCGCTTTCACCTATCCAGGCAATTTGAAATCTATTGGGA 3966
Db 3901 CCCATGCTGACCTGGAGCACCTCGCGCTTTCACCTATCCAGGCAATTTGAAATCTATTGGGA 3960
QY 3967 GATGAAGGAAAACCTCTGTTTGGAGCACCTTCAAAATAGCAGCATTAATGGTCTGAAAGCA 4026
Db 3961 GATGAAGGAAAACCTCTGTTTGGAGCACCTTCAAAATAGCAGCATTAATGGTCTGAAAGCA 4020
QY 4027 TTAATGCAAGTTGCAAGTTCACAGAGGATTACCTGTCCTCAGTCTGATACAGAAACAT 4086
Db 4021 TTAATGCAAGTTGCAAGTTCACAGAGGATTACCTGTCCTCAGTCTGATACAGAAACAT 4080
QY 4087 CTGGTTCGTTCTATCAGTTCTTCTCCTAACATAAATAAGAGATACACCATGCCCTT 4146
Db 4081 CTGGTTCGTTCTATCAGTTCTTCTCCTAACATAAATAAGAGATACACCATGCCCTT 4140
QY 4147 CTGCTATAGATCTGTTTCAATGTTTGGTGGTGTGTGTAGCATCCCATCCTTGAT 4206
Db 4141 CTGCTATAGATCTGTTTCAATGTTTGGTGGTGTGTGTAGCATCCCATCCTTGAT 4200
QY 4207 TGGGATGACCCCTGTTGATCTGAGCTTCTCAGTAGTCTTCTCCTATACACCTTAT 4266
Db 4201 TGGGATGACCCCTGTTGATCTGAGCTTCTCAGTAGTCTTCTCCTATACACCTTAT 4260
QY 4267 CTCTTCCATTTGATCACCATTGACACATGCTTTCAGATACACTACTACTAGACAGGC 4326
Db 4261 CTCTTCCATTTGATCACCATTGACACATGCTTTCAGATACACTACTACTAGACAGGC 4320
QY 4327 CTACCCCTTGCTCAGTTCAAGAAGACAGTGAAGAGGCTCATTCGCCATCTTCTTCTT 4386
Db 4321 CTACCCCTTGCTCAGTTCAAGAAGACAGTGAAGAGGCTCATTCGCCATCTTCTTCTT 4380
QY 4387 GCAGAAATTTCTCAATATACAGTGGCTCCATTTGGTGTGATATTCCTGGCTGGTATTG 4446
Db 4381 GCAGAAATTTCTCAATATACAGTGGCTCCATTTGGTGTGATATTCCTGGCTGGTATTG 4440
QY 4447 TGGGTCTCAGTGAAGAATGGCATACCCCTTATCTTCGCTGTGCTGATTTGTTTCCAC 4506
Db 4441 TGGGTCTCAGTGAAGAATGGCATACCCCTTATCTTCGCTGTGCTGATTTGTTTCCAC 4500
QY 4507 TATTTACTTGGGTAACTCCGCTGAGGAATGCATACCAATTTCTCAGAAGGAGAGTAC 4566
Db 4501 TATTTACTTGGGTAACTCCGCTGAGGAATGCATACCAATTTCTCAGAAGGAGAGTAC 4560
QY 4567 AGTGCACTCTGAGCTATCTTATCTTACCTACAAATTTGTCCTGCTCTTCCAGGAATAT 4626
Db 4561 AGTGCACTCTGAGCTATCTTATCTTACCTACAAATTTGTCCTGCTCTTCCAGGAATAT 4620
QY 4627 TGGGATCTGTAAGCCCTTGTCTCAGAGGCGGTGTGCAGATCCTGCCCTTACTAACTGT 4686
Db 4621 TGGGATCTGTAAGCCCTTGTCTCAGAGGCGGTGTGCAGATCCTGCCCTTACTAACTGT 4680
QY 4687 TTGAAGCAAAAAACACCGTGTGCTAGCTACCTAGAAAAGAAATAGTTGATAGAGCTT 4746
Db 4681 TTGAAGCAAAAAACACCGTGTGCTAGCTACCTAGAAAAGAAATAGTTGATAGAGCTT 4740
QY 4747 CCTGATGACTATAGCTGCTCTGTAATCAAGCTTCTCATTTTCAGGTGCGCCACGCTGCA 4806
|||||
```

```
Db 4741 CCTGATGACTATAGCTGCCTCCTGAAATCAAGCTTCTCATTTTCAGGTGCCACGGTCTGCA 4800
QY 4807 GATGATGAGGAAAAGCATCTGTCTCTGCTGCTCTTCTCTGGGGCTATACTATGTTCTCAG 4866
Db 4801 GATGATGAGGAAAAGCATCTGTCTCTGCTGCTCTTCTCTGGGGCTATACTATGTTCTCAG 4860
QY 4867 AACATTTTCTGCCAGGAAATTTGTGAACGGGGAAGAGTTGGAGCTTGATTTTTCACGCA 4926
Db 4861 AACATTTTCTGCCAGGAAATTTGTGAACGGGGAAGAGTTGGAGCTTGATTTTTCACGCA 4920
QY 4927 CTTTCACTGTGAGCGGAGTCTGCAATTTTCTTAAAAATCAGAGAATCGGAGTGTCTCTG 4986
Db 4921 CTTTCACTGTGAGCGGAGTCTGCAATTTTCTTAAAAATCAGAGAATCGGAGTGTCTCTG 4980
QY 4987 GTTGAAGGTAAAGCCAGAGGCTGTGCCTATCCAGCTCCTTACTTGGATGAATATGAGAA 5046
Db 4981 GTTGAAGGTAAAGCCAGAGGCTGTGCCTATCCAGCTCCTTACTTGGATGAATATGAGAA 5040
QY 5047 ACAGACCTTGCCTTGAAGAGGGCAACCCCTTCAATTTATCTCTGAGGGGTATCGGAAG 5106
Db 5041 ACAGACCTTGCCTTGAAGAGGGCAACCCCTTCAATTT-TCTCTGAGCGGTA-CGGAAG 5098
QY 5107 CTCCATTTGGTCTGGCAACAACACTTGCATTA 5137
Db 5099 CTCCA-TTGTNCGGCAACACACTTGCATAA 5128

RESULT 3
AR030784 LOCUS AR030784 6395 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 1 from patent US 5861312.
ACCESSION AR030784
VERSION AR030784.1 GI:5943998
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 6395)
AUTHORS Varshavsky, A. and Kwon, Y. Tae.
TITLE Nucleic acid encoding mammalian UBR1
JOURNAL Patent: US 5861312-A 1 19-JAN-1999;
FEATURES
source
location/Qualifiers
BASE COUNT 1802 a 1376 c 1519 g 1698 t
ORIGIN

Query Match 78.5%; Score 4085.6; DB 6; Length 6395;
Best Local Similarity 86.6%; Pred. No. 0;
Matches 4567; Conservative 9; Mismatches 629; Indels 69; Gaps 4;

QY 1 ATGCGCGACGAGGAGGCTGGAGGTACTGAGAGGATGGAATCAGCGCGGAGTTACCCCGAG 60
Db 115 ATGCGCGAGGAAGAGATGGAGCGCGCGGAGGAGTGGACGTGAGCGCGGAGGCTCCCTG 174
QY 61 ACCCCTCAGGCTCTGGCATCTTGGTGGGATCAGCAAGTTGATTTTATCTACTGTTCTTG 120
Db 175 GCCCGGAGCGCGCGCATCTGTTGGGATCAGCAAGTTGATTTCTACTGCTTCTTA 234
QY 121 CATCATTTGGCACAAATTTGGTCCGCAAGAAATTTACTTGTCTGAAATGGACCCGAGTGGAA 180
Db 235 CATCATTTGGCACAAATTTAGTCCGCAAGAAATTTATTTGCTGAGATGGACCCGAGTGGAA 294
QY 181 AAGCAGAGGAGGAAGTGTACAAATGTCAATATCTACTCCACTCCGATGGATGTTATTTGA 240
Db 295 AAGCAAGAAGAGGTGTACAGATGTCAATACTCCTCTTGGAGTGGTACTTATTTGA 354
QY 241 GAAGATCCAGATTTTCTTAGAGAAATTTGAAGCACAGTGGAGCATTTTCAGCTTTGTGG 300
Db 355 GAGGATCCGATATTTGCTTTAGAGAAATTTAAACACAGTGGAGCGTTCAGTGTGTGG 414
QY 301 AGGTTTTTCAAAAAGTGGAGAGACAACCTTATTTCTTTCAGGGATTTGTGCAATTTGATCAACA 360
|||||
```


QY 2476 ACCGACATAGCAGGCTGAACATATGTCAGAGAAAAGGAGAAAACAGAAACAAAGAT 2535
DB 2635 ACACAGATAGCAAGGCTGAACATATGTCAGAGAAAAGGAGAAAACAAAGAT 2694
QY 2536 GAAGCATTTGCCGACACACACCTCTGTAATTTCTGCCCTGCTTTGACGCGGCAATTAAC 2595
DB 2695 GAAGCATTTGCCGCGCCACCTCTCCAGAGTTCTGCCCTGCTTTGACGCGGCAATTAAC 2754
QY 2596 CTTCTCACTGTGATATCATATGATATCTCAGGACCGTATTTGAGCGGCAATTAAC 2655
DB 2755 CTGCTCAGCTGTGATGTATGATATATATCTCAGGACCGTATTTGAGCGGCAATTAAC 2814
QY 2656 ACAGATTTCAACTTTGGACGAGGAGTCTCCAAATGCTTTTATATCTGCGCATTTG 2715
DB 2815 ACGAGCTCTAATCTGTGGACAGAGGAGTCTCAGATGCGTTCCTATATATTGGCACTG 2874
QY 2716 GGTTTTACTAGAAGAGAACACAGCTTCAAAAAGCTCCTGAAGAAGAGTAAATTTGAC 2775
DB 2875 GGCTTCTGGAAGAGAACGACAGCTTCAGAAGCTCCTGAAGAGAGTGGCTTTTGAC 2934
QY 2776 TTTTATCATAGGCTTCAAGATTTGGAAGTTTCAGCCAT-----GAATATACAAATG 2826
DB 2935 TTTTACCAATAAGCTTCAAGATTTGGAAGTTTCAGCCATGAATGCTCAGAAATATACAATG 2994
QY 2827 CTTTGGAAAACTCAAGAAATTTCCCGAGTTAGAGGCCAGAGGACATGATAACGTGG 2886
DB 2995 CTCCTTGGAAAGACTCAAGAGAAATTTCCCGCAATTTAGAAGGCCAGAGGACATGATAACATGG 3054
QY 2887 ATACTTCAGATGTTTTCACACAGTGAAGCGATTAAGAGAAAATCTGTTTAAATTTAGCA 2946
DB 3055 ATACTCCAGATGTTTTCACACAGTGAAGCGATTAAGAGAAAATCTGTTTAAATTTAGCA 3114
QY 2947 ACCACATCAGGATTCGAATCTATTAGAATGATGAGATTAATCTCATGATAAAGAAAAGCA 3006
DB 3115 ACCACTTCAGGACTGGAGTGCAATTAAGAGTGAAGATTAATCTCATGATAAAGAAAAGCA 3174
QY 3007 GAACGAAAAGAAAGCTGAAGCTGCTAGGCTACATGCGCAGAGATCATGGCTCAGATG 3066
DB 3175 GAACGAAAAGAAAGCTGAGGCGCTGAGGCTTCATGCGCAGAGATCATGGCCAGATG 3234
QY 3067 TCTGCTTACAGAAAATCTCATTTAACTCATAACTCATGATGATGACATATACATCAGAA 3126
DB 3235 TCTGCTTACAGAAAATCTCATTTAACTCATGAACTCATGATGATGATGATGATGATGATG 3294
QY 3127 ATGCTTGGAAAGAAAGATTCATTTATGGAGAGAGAGACCCAGCAGTCACTGACTAC 3186
DB 3295 GTAACAGGGAAGAAAGACTCCATTTATGGAGAGAGAGACCCAGCAGTCACTGAGGCC 3354
QY 3187 TCTAGAAATTTGCTTTGGGCTTAACCGGGTCCATCTGTTACTGAAAGAGGCTGCTGACG 3246
DB 3355 TCTAGAAATTTGCTTTGGGCTTAACCGGGTCCATCTGTTACTGAAAGAGGAGGCTGACG 3414
QY 3247 TGCATCTTTGCCAAGAAAGAGGAGTCAAAATAGAAAATTAATGCCATGGTATTATCG 3306
DB 3415 TGCATCTCTGCCAAGAAAGAGGAGTCAAAATAGAAAATTAATGCCATGGTATTATGCA 3474
QY 3307 GCCTGTGTCGAAATTTCTACTGCTTAAACCCAGCAGGAGGAAAACCCATAGAACTCTCA 3366
DB 3475 GCATGTGTCGAAATTTCTACTGCTTAAACCCAGCAGGAGGAAAACCCATAGAACTCTCA 3534
QY 3367 GGAGAAGCCCTAGACCCACTTTTTCATGATCCAGACTTGGCATTATGGAATTTATACAGGA 3426
DB 3535 GGGGAAACACTGGACCTCTTTTTCATGATCCAGACTTGGCATTATGGAATTTATACAGGA 3594
QY 3427 AGCTGTGTCATGTAATGACGAGTGTGCTGGCAGAAATTTTGAAGCTGTACAGCTG 3486
DB 3595 AGCTGTGTCATGTAATGATGATGATGCTGCTGGCAGAAATTTTGAAGCTGTACAGCTG 3654
QY 3487 AGCTCTCAGCAGCGATTCATGTTGACCTTTTGGCTTGGAGTGGAGAAATTTCTTTGC 3546
DB 3655 AGCTCGCAGCAGCGATTCATGTTGACCTTTTGGCTTGGAGTGGAGAAATTTCTTTGC 3714
QY 3547 CCTCTTTGCAAACTCTGTGCAATACTGTGATCCCTTATCTCTTTGCAAACTCAAAAG 3606

DB 3715 CCGCTCTCAAGTCTCTCTGCAACACTGTCTATCCCATCATCCCTTTGACGCGGAGAAG 3774
QY 3607 ATAAACAGTGAAGATGCAGATGCTCTGCTCAACTTTTTCACCTGCGCAGGTCGATACAG 3666
DB 3775 ATCAACAGTGAAGATGCGGAGGCTCTGCTCAACTTTTTCACCTGCGCAGGTCGATACAG 3834
QY 3667 ACTGTTCTGGCCAGAAATATCAGGTTTATATATAAGACATGCTAAAGGAGAAAAC--CA 3723
DB 3835 ACTGCTCTGCGAGAAATATCGGGTTTATATATAAGACATGCTAAAGGAGAAAACCCAGCA 3894
QY 3724 ATTCTATTTTCTTTTAAATCAAGGAATGGAGATTTCTACTTTTGGAGTTTCCATTCCTG 3783
DB 3895 GTTCTCTGCTTTTAAATCAAGGAATGGGGAATCAACTTTTGAATTTTCAATTCATTCCTCTG 3954
QY 3784 AGTTTGGCGTGTGAGCTTTCGATTAATTAATTAAGATGATCAAGGAAATGGTATTCTC 3843
DB 3955 AGTTTGGAGTTCAGTCTTCGTTGAATATTAATTAAGATGATCAAGGAAATGGTATTCTC 4014
QY 3844 TTTGCCACAACAATTTATAGAAATGGATTTGAAAGTCCCACTGATGAAAGGGATCCTCGA 3903
DB 4015 TTCGCCACAACAATTTACAGAAATGGCTGAAAGTGCCTCTGATGAACATAGACCCACGA 4074
QY 3904 GTCCCATGCTGACCTGGAGCACCTGGCGTTTCTACTATCCAGGCAATTTGAAATCTATTG 3963
DB 4075 GTGCCCATGATGACCTGGAGCACCTGTGCTTCCAGCATCCAGGCAATTCGAAACCTGTTG 4134
QY 3964 GGAGATGAAGAAACACCTCTGTTTGGAGCACCTTCAAAATAGGAGCATTAATGGTCTGAAA 4023
DB 4135 GGAGATGAAGAAACACCTCTATTTGGAGCACCTTCAAAATAGAGCATAGCGCTCTGAAAG 4194
QY 4024 GCATTAATGCAAGTTTGCAGTTTGCACAGAGGATTAAGTCTGCTCAGTCTATATACAGAAA 4083
DB 4195 GCGCTAATGCAAGTTTGCAGTTTGCACAGAGGCTTACTGCGCTCAGTCTCTGATACACAA 4254
QY 4084 CATCTGTTCTGCTTCTATCATGTTTCTTCTAATCAATATAAATAGAGATACACATGCT 4143
DB 4255 CATCTGCTCGCTCTCTGCTGATTTCTTCTTAACTGCAATCAGAAAATACACAGGC 4314
QY 4144 CTTCTGCTATAGATCTGTTTCTGCTGTTTGGTGGTGTGTTAGCATTTCCATCTCTG 4203
DB 4315 CTTCTGCTGTTGATCTCTTCCATGTTCTGCTGGGCGAGTCTTAGCTTTCCATCTCTG 4374
QY 4204 TATTGGATGACCTGTTGATCTGCGAGCTTCTTCAAGTATGTTCTTCTTATACCACTT 4263
DB 4375 TATTGGATGACCGTGTGATCTGCGAGCGTGCACCTTACTTCTTATATACCACTT 4434
QY 4264 TATCTTCTCATTTGATCACCATGCGACACATGCTTCAAGTACTACTTACGTTAGACA-- 4321
DB 4435 TATCTTCTCATCTGATCACCATGCGGACATGCTTCAAGTACTCTTACAAACAGATACA 4494
QY 4322 -----CAGGCTTACCCCTTGTCTGAGTTTCAAGAGACAGTGAAGGCTCATTC 4371
DB 4495 GATCTGCTCCAGGCGCGCGCTGCTGAGGTTGAAGGATAGTGAAGGCTGCTGCT 4554
QY 4372 GCATCTTCTTCTTTGAGAAATTTCTCAATATACAAGTGGCTCATTTGGTGTGATATT 4431
DB 4555 GCATCTGCTTCTTTGAGAAATTTCTGCGAGCACACAGCGGCTCCTGAGGTCGCTGCT 4614
QY 4432 CCTGGCTGGTATTTTGGGTTCTCACTGAAGAAATGGCATCACCCCTTATCTTCTGCTGCT 4491
DB 4615 CCGGCTGGTACCTGTGCTCTCCCTGAGGAACGGCATCACCCCTTACCTCCGCTGCTGCT 4674
QY 4492 GCATTTGTTTTCACATTTTACTTTGGGTTACTCCGCTGAGGAAGTGAATACCAATTC 4551
DB 4675 GCATGCTTTTTCACATTTTACTTTGGAGTAGTCTCCGCTGAGGAAGTGTGTTCCCAATTC 4734
QY 4552 GCAGAAGAGAGTACAGTGCATCTGTAGCTATCTATCTTTACCTACAAATTTGTTCCG 4611
DB 4735 GCTGAAGAGAAATTCAGTGCATCTGTAAGTATCTATCTTTTACCACAAATTTGTTCCG 4794
QY 4612 CTCTTCCAGGAATATTGGGATAGTGAAGGCCCTTGTCTCCAGAGGCGGTGTCAGATCT 4671
DB 4794 GCTGAAGAGAAATTTGGGATAGTGAAGGCCCTTGTCTCCAGAGGCGGTGTCAGATCT 4671

4795	CTTTTCCAGGAATATTGGGATACCATTAAGGCCCTTACTACAGAGGTGGTGGAGATCCT	4854
4672	GCCTTACTAACTGTTTGAAGCAAAAAACACCGTGCAGGTACCCCTAGAAAAAGAAAT	4731
4855	GCCTTACTCAAGTCTTTGAAGCAGAAAGTGTGTGGTCAAGTACCCCTAGAAAAAGAAAT	4914
4732	AGTTTGATAGAGCTTCCCTGATGACTATAGTGCCTCCTGAAATCAAGCTTCTCATTTACAGG	4791
4915	AGTTTGATAGAGCTTCCCTGAGGACTACAGCTGTCTTAAATCAGGCTTCTCATTTAGG	4974
4792	TGCCCCAGGCTGTCAGATGATAGCGAAAGCATCCTGTCTCTGCTCTTTCTGTGGGGCT	4851
4975	TGTCACGGTCTGCAGATGATGAGCGAAAGCATCCTGTCTCTGCTCTTTCTGTGGGGCC	5034
4852	ATACTATGTTCTCAGAACATTTGCTGCCAGGAAATTTGTAAGCGGGAAGAGGTTGGAGCT	4911
5035	ATCCTGTGTTCTCAGAACATCTGTTGCCAAGAAATAGTGAATGGGGAAGAGGTTGGAGCG	5094
4912	TGCATTTTTCACGCACTTCACCTGTGGAGCGGAGTCTGCATTTTCTTAAAAATCAGAGAA	4971
5095	TGCGTTTTCATCGCTTCATTTGTGTGTGGAGTCTGCATTTTCTTAAAAATCCGAGAA	5154
4972	TGCCGAGTGGTCTGTTGAGGTAAGCCAGAGGCTGTGCCTATCCAGCTCCTTACTTGG	5031
5155	TGCAGGTTGGTCTGTTGGAAGGAAAGCCAGAGGCTGTGCCTATCCAGCCCTTACTTGG	5214
5032	GATGAATATGGAGAAACAGACCTTGCCTGAAGAGGGCAACCCCTTCATTTATCTCGT	5091
5215	GATGAATATGGAGAAACAGACCCAGGCTTAAGAGAGGAAACCCACTTCATTTATCTCGG	5274
5092	GAGCGGTATGGAGGCTCCATTTGGTCTGGCAACACACTGCAATTTAGAGAAGATTGCT	5151
5275	GAGCGGTATCGGAAGCTGCAATTTGGTCTGGCAACAGCACTGCAATTTAGAGAAGATTGCT	5334
5152	AGAGCCAGAGACTAATCAGATGTTATTGGATTCAACTGGCAGTTACTGTGA	5205
5335	CGGAGCCAGAGACTAATCAGATGCTATTGGATTAACTGGCAGTTACTCTGA	5388
RESULT 4		
AR121463		
LOCUS	6395 bp	DNA
DEFINITION	Sequence 1 from patent US 6159732.	linear
ACCESSION	AR121463	PAT 16-MAY-2001
VERSION	AR121463.1	
KEYWORDS	GI:14105039	
SOURCE	Unknown.	
ORGANISM	Unknown.	
REFERENCE	1 (bases 1 to 6395)	
AUTHORS	Varshavsky, A. and Kwon, Y. Tae.	
TITLE	Nucleic acid encoding mammalian Ubr1	
JOURNAL	Patent: US 6159732-A 12-DEC-2000;	
FEATURES	Location/Qualifiers	
source	1..6395	
BASE COUNT	1802 a 1376 c 1519 g 1698 t	
ORIGIN		
Query Match	78.5%; Score 4085.6; DB 6; Length 6395;	
Best Local Similarity	86.6%; Pred. No. 0;	
Matches 4567; Conservative	9; Mismatches 629; Indels 69; Gaps 4;	
QY	1	ATGGCGGACGAGGAGGCTGGAGGTACTGAGAGATGGAATATCAGCGGGAGTTACCCCG 60
Db	115	ATGGCGGACGAGGAGATGGAGCGCGCGAGAGGATGGACGTGACCGCGGAGCCCTCCCG 174
QY	61	ACCCTCAGCGCTGGCATCTGGTGGATCAGCAAGTTGATTTTATCTGCTTCTTCTG 120
Db	175	GCCCCGCGGCGCGGATCGTGGTGGATCAGCAAGTTGATTTTCTATCTGCTTCTTA 234
QY	121	CATCATTTGGCACAATTTGGTCCAGAAATTTACTTTTCTGCTGAAATGGACCCAGACTTGGAA 180
Db		

QY	1216	ATGTTTACTGTTCTTACTCTGCTCGACATCTTATTGAAGAGCAGAAATGTTATCTCTGTC	1275	QY	2296	ATGCCACACAGTGGCCATTGGCCAAAATTTACCTGAGATGAAATTAATGAACCTGGCTTA	2355
DB	1375	ATGCTCACCGTCCCGACCTTGGCCGGCACTTATTGAAGAGCAGAAATGTTATCTCTGTC	1434	DB	2455	ATGCCACACAGTGGCCATTGGCCAAAATTTACCTGAGATGAAATTAATGAACCTGGCTTA	2514
QY	1276	ATTACTGAAACTCTCTCTAGAACTTTTACCTGAGTACTTGGACAGGAAACAATAAATTCAC	1335	QY	2356	GAGAAATGCTATAAACAAGTGGCCACATTTTAAAGAACCCAGGTGATATCAGGCAATGAGTT	2415
DB	1435	ATTACTGAAACGCTCTCTAGAACTTTTACCTGAGTACTTGGACAGGAAACAATAAATTCAC	1494	DB	2515	GAGAAATGCTATAAACAAGTGGCCACATTTTAAAGAACCCAGGTGATATCAGGCAATGAGTT	2574
QY	1336	TTCAGAGGTTATAGCCAGACAAAATTTGGGAAGATATATGCGAGTAAATGTCGACCTAAAG	1395	QY	2416	TATGAACATAAAGATGAATCCTGAAAGACCTCAATATGATGTTACTTTTATCATTTACTTCCAAA	2475
DB	1495	TTCAGAGGTTATAGCCAGACAAAATTTGGGAAGATATATGCGAGTAAATGTCGACCTAAAG	1554	DB	2575	TATGAATTTGAAGATGAATCCTGAAAGACCTCAATATGATGTTACTTTTATCATTTACTTCCAAA	2634
QY	1396	TATATCCCTGATAGCAACCCACATATATGGACAGAAAGATTAAAGATGAGTTCCTTGAA	1455	QY	2476	ACCCAGCATAGCAGGCTGACATATGCAGAGAAAGAGGAGAAACAAAGAAACAAAGAT	2535
DB	1555	TATATCCCTGATAGCAACCCACATATATGGACAGAAAGATTAAAGATGAGTTCCTTGAA	1614	DB	2635	ACACAGCATAGCAGGCTGACATATGCAGAGAAAGAGGAGAAACAAAGAAACAAAGAT	2694
QY	1456	GGTTTTCGATCTTTTGAAGATCTTACCTGATGCGAGGAAATGGAAGAAATCCGAAGA	1515	QY	2536	GAAGCATTTGCCGCCACCACCTCCCTGAAATTTCTGCCCTGCTTTCAGCAAAAGTGAATAC	2595
DB	1615	GGTTTTCGATCTTTTGAAGATCTTACCTGATGCGAGGAAATGGAAGAAATCCGAAGA	1674	DB	2695	GAAGCATTTGCCGCCACCACCTCCCTGAAATTTCTGCCCTGCTTTCAGCAAAAGTGAATAC	2754
QY	1516	CAGTTTGGCAACACATTTGAAGTGGATCCTGATTTGGAGGCTGCCATTTGCTATACAGATG	1575	QY	2596	CTTCTCAACTGTGATATCATGATGATCATTTCTCAGGACCCGTTATTTGAGCGGCAATAGAC	2655
DB	1675	CAAGTTGGCAACACATTTGAAGTGGATCCTGATTTGGAGGCTGCCATTTGCTATACAGATG	1734	DB	2755	CTGCTCAGCTGTGATGTTATGATATACATCCTCAGGACCATCTTTGAGCGGCAATAGAC	2814
QY	1576	CAATTTGAAGATATTTTACTCATGTTTCCAGAGTGGTGTGCTGTGATGAAGAACTCTTA	1635	QY	2656	ACAGATTTCACTTTGTGGACCGGAGGATGCTCCAAATGGCTTTTTCATATTTCTGGCATG	2715
DB	1735	CAACTTAAAGATATTTTGTCTCATGTTTCCAGAGTGGTGTGCTGTGATGAAGAACTCTTA	1794	DB	2815	ACGGAGTCTAATCTGTGGACAGAGGATGCTGCAGATGGCTTCCATATATTTGACATG	2874
QY	1636	CTTGTGGCTTATAAGAAATGTCACAAAGCTGTGATCAGGTGAGTACCAGTTTCATATCT	1695	QY	2716	GGTTTACTAGAAGAGAGCAACAGCTTCAAAAAGCTCCTGAAGAGAAAGTAACTTTGAC	2775
DB	1795	CTTGTGGCTTATAAGAAATGTCACAAAGCTGTGATCAGGTGAGTACCAGTTTCATATCT	1854	DB	2875	GGCTTGTGAGAGAGAGCAGCAGCTTCAGAAAGCTCCTGAAGAGAAAGTGGCTTTGAC	2934
QY	1696	AGTAGCAAGACAGTAGTACAAATCGTGTGACATAGTTTGGAAACAAAGTCCCTACAGATG	1755	QY	2776	TTTATATCAAGGCTTCAAGATTTGGGAAGTTTCAGCCAT - - - - - GAATATACAAATG	2826
DB	1855	AGTACCAAGACAGTAGTACAAATCGTGTGACATAGTTTGGAAACAAAGTCCCTACAGATG	1914	DB	2935	TTTATATCAAGGCTTCAAGATTTGGGAAGTTTCAGCCATGATGCTCAGAAATATACAAATG	2994
QY	1756	TCTGAGGATCTTGAAGATATCATCTGCCACTCTCTAGGACCCCTGCTGGTCTTCATGTA	1815	QY	2827	CTTTTGGAAACATCAAGGAATTTCCCGAGTTTGAAGGCCAGAGGACATGATTAACGTGG	2886
DB	1915	TCTGAGGACCTTGTGAAGATATACCTGCCACTCTCTAGAACACTTCTGCTGGTCTTCATGTA	1974	DB	2995	CTCTTGGAAAGACTCAAGGAATCCCCCAATTTAGAAGGCCAGAGGACATGATTAACATGG	3054
QY	1816	CGTTTAAAGCAGCTGGTGTCTTTCAAGACTGCATGAATTTGCTTTTGGAGACTTT	1875	QY	2887	ATACTTCAGATGTTTGACAGTGAAGCTTAAGAGAAATCTCTGTTTATTTCTAGCA	2946
DB	1975	CGTTTAAAGCAGCTGGTGTCTTTCAAGACTGCATGAATTTGCTTTTGGAGACTTT	2034	DB	3055	ATACTTCAGATGTTTGACAGTGAAGCTTAAGAGAAATCTCTGTTTATTTCTAGTGCA	3114
QY	1876	CAAGTAGAGGTACTAGTGAATATCTTTACCTGTGCTGGTGTGGTTCGCCAGAGTTGTT	1935	QY	2947	ACCACATCAGGATCGGAATCTTATTAAGAAATGATGAGATTTACTCATGATTAAGAAAGCA	3006
DB	2035	CAAGTAGAGGTCTGCTGGAGTACCGCTGCCCTGGCTGGTCTGGTCTGAGTTGTT	2094	DB	3115	ACCACATCAGGATCGGAATGCAATTAAGAGTGAAGATTTACTCATGATTAAGAAAGCA	3174
QY	1936	GCTGAGATGTGGCGAAGAAATGGACTGTCTCTTATTAGCCAGGTGTTTATTACCAAGAT	1995	QY	3007	GAACGAAAGAAAGCTGAAGCTGCTAGGCTACATGCCAGAGATCATGGCTCAGATG	3066
DB	2095	GCTGAGATGTGGCGAAGAAATGGACTGTCTCTTATTAGCCAGGTGTTTATTACCAAGAT	2154	DB	3175	GAACGAAAGAAAGCTGAAGCTGCTAGGCTTTCATGCCAGAAAGATCATGGCCAGATG	3234
QY	1996	GTTAAGTGCAGAGAAAGATGATGATTAAGATATCATGCTTCAGATTTGGTGATCT	2055	QY	3067	TCCTGCCTTACAGAAACTTCAATGAACTCATAACTCATGATGATGATGATGATGATGATG	3126
DB	2155	GTTAAGTGCAGAGAAAGATGATGATTAAGATATCATGCTTCAGATTTGGTGATCT	2214	DB	3235	TCCTGCCTTACAGAAACTTCAATGAACTCATAACTCATGATGATGATGATGATGATGATG	3294
QY	2056	TTAATGGATCCCAATAGTCTTGTACTGGTACTTCAGAGGTATGAACTTCCGAGGCT	2115	QY	3127	ATGCCTGGGAAAGAAAGATTCATTTATGGAGGAGAGAGCAGCCAGTCACTGATGATG	3186
DB	2215	ATAATGGATCCCAATAGTCTTGTACTGGTACTTCAGAGGTATGAACTTCCGAGGCT	2274	DB	3295	GTAACAGGAGAGAGAGTCCATTTATGGAGGAGAGAGAGCAGTCACTGAGTGGGCT	3354
QY	2116	TTTAAAGACCATATCTACAAAGACAGGATTTGATTAAGCAATATATACACTTAATA	2175	QY	3187	TCTAGAAATGCTTTGGGTCCTTAAACGGGTCCTTGTGTTACTGAAAGAGAGGTGCTGACG	3246
DB	2275	TTTAAAGACCATATCTACAAAGACAGGATTTGATTAAGCAATATATACACTTAATA	2334	DB	3355	TCTAGAAATGCTTTGGGTCCTTAAACGGGTCCTTGTGTTACTGAAAGAGAGGTGCTGACG	3414
QY	2176	GAAGAAATGCTTCAGTCTCATATATATTTGGGTGAGCGTTATGTAACCTGGAGTGGGA	2235	QY	3247	TGCAATCCTTTGGGTCCTTAAACGGGTCCTTGTGTTACTGAAAGAGAGGTGCTGACG	3306
DB	2335	GAAGAAATGCTTCAGTCTCATATATATTTGGGTGAGCGTTATGTAACCTGGAGTGGGA	2394	DB	3415	TGCAATCCTTTGGGTCCTTAAACGGGTCCTTGTGTTACTGAAAGAGAGGTGCTGACG	3474
QY	2236	AATGTGACCAAGAGAGGTTCATATGAGAGAAATCATTCATGCTTGTGATTTGACCT	2295	QY	3307	GCCTGTGTCAGAAATCTACTGCTTAAACGGGTCCTTAAACGGGTCCTTGTGTTACTGAA	3366
DB	2395	AATGTGACCAAGAGAGGTTCATATGAGAGAAATCATTCATGCTTGTGATTTGACCT	2454	DB	3475	GCATGCTGTCAGAAATCTACTGCTTAAACGGGTCCTTAAACGGGTCCTTGTGTTACTGAA	3534
				QY	3367	GGAGAGCCCTAGACCCACTTTTTCATGATCCAGACTTGGCATATGGAACCTTATACAGGA	3426

TITLE Direct Submission
JOURNAL Submitted (24-APR-1998) Division of Biology, 147-75, California
Institute of Technology, 1200 E. California Bl., Pasadena, CA
91125, USA

FEATURES
source Location/Qualifiers
1..6395
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/chromosome="2"
/cell_line="MEL-C19"
1..6395
/gene="Ubr1"
1..114
/gene="Ubr1"
115..5388
/gene="Ubr1"
/note="UBR1p; recognition component of the N-end rule
pathway"
/codon_start=1
/product="ubiquitin-protein ligase E3-alpha"
/protein_id="AAC0165.1"
/db_xref="GI:3170887"
/translation="MADEMDGAERMDVPEPLAPORPASWMDQVDVYAFILHLHA
QLVPEYFAEMDPDLKQESVQMSILTPLEWLFGEEDPDICLEKLEKHSAPQLCKGY
FSGETYSRCRAIDPTVCLMDCFSQSSVKNHRYKMHSTGGGDCDEGDEAWKTLQ
PCVDHEPGRAGTKESLHCPLENEVIAQRFPPIVYKIVEMTIWECEKELPELQ
IREKNRYCVLFNDSHSDVHYVLSQVADCELAELAHHTAIDKGRRAVKGAVY
ATCOEAKDIKSHSNVSOPLHVEVLSQVADCELAELAHHTAIDKGRRAVKGAVY
FCQACLVPEPSSNCLISRLMLDAKLYKGBKILHELIFSSFFEMEMKYLAMEF
VXYQLQKEYISDHERSISITALSVMQLVPTLRLHLEQNVLSVITETLLEVL
EYLRNKNFQSGYSDKLGRIYVYICDLKYLISKPIVWTERLRAQFLEGFSEFLKI
LTCOMGMEIRFQGHIEVDPDEAAIAIQMLKNILMFQEWACDEDLVAYKE
CHKAVRSTNFMSTKTVVQCGHSLKTSKVSDELVSILPLSLTLAGLHVLSR
LGAISRLFEVDFDSQVEVYPLRCLVLAQVVAEMWRNGLSLISOVFYQDVK
CREMYDKDIIMLOIGASIMDPNKLFLVLYVELDAENKTLSTQDQLKOYNLIL
EMQLVLIYIGERYVPGVNTREEVIMREITHLILCIEPMPHSAIARNPENNET
GLENLVINKATPEKPGVSGHVTLEKSDKPMFYFIYHYSKTOHSAEHMQKRRKQ
ENKDRALPPPPPEFCFAPFVNVNLSQDMYILRTIFERAVDTESNLWTEGLQMA
PHILALGEEKQQLQKAPKEEVAFDFYHKAISLGSSAMNAQNIOMLLERLKGIPOLE
GOKDMITLWLOFDTVKRLREKSLVATTSGLECIKSEETHDKBAERKRAEAR
LHROKIMAOASLQKNIETHLMYDNTSEVTKEDSIMEESTSAVSEARLQPK
RGPATKEVILICQERQEVKLENNAMVLSNVCQKSTALTOHGRKPDVHLETDP
LWMDPLAGTTCGGHVAWQKFEVOLSQQRHVDLSEGEYLCPLCK
SLCNTVPIPILOPKINSENAEALQTLTARWQTLVARSYNIHAKGAPAVP
VLENGMGDSTFEHSILSGVSSVYKSNIKEMVLFATTIYRIGLKVPPDELDPR
VPMWTSTCAFTIATENLLGDEKPLFGALONROHSLKALMOFAVAORATCPVLI
KHHLARLLSVLPNLOSENTPGLSLVDLFLVGLVLAFLPSLWDDTVLOPSLSS
YNHLXILFHTMAHMLQILITDLSLSPPLAGEDESEERACASAPFVFSQHTDG
LFGCGAPGWLMLSLRNGITPYLRCAALLFHYLLGVAPPEELFANSABEFSALCSYL
SLPTNLLFLFYEDTIRFLQRCWGDPAALLSKOKSAVVYPRKNSLIPELSDYS
CLLNQASHPCRSADDERKHPVLCIFCGAILCSQNICQEIIVNGEVGACVPHALHC
GAGVCIFLIRRCVYLVGKARGCAYPAPYLDEYGETDPGLKRGPNPLHLSRERYKL
HVAWQHCILIEETARSQETNQMLFGFNWQLL"
5389..6395
/gene="Ubr1"
polya_site 5429..5434
/gene="Ubr1"
BASE COUNT 1802 a 1376 c 1519 g 1698 t
ORIGIN

Query Match 78.5%; Score 4085.6; DB 10; Length 6395;
Best Local Similarity 86.6%; Pred. No. 0;
Matches 4567; Conservative 9; Mismatches 629; Indels 69; Gaps 4;

QY 1 ATGGCGGACGAGAGCTGGAGTACTGAGAGGATGAATACGCGCGGAGTACCCCG 60
DB 115 ATGGCGGACGAGAGATGACGCGCGGAGAGGATGAGCTAGCGCGGAGCTCCCGT 174
QY 61 ACCCTCAGCGCTGTCATCTGGTGGGATCAGCAAGTTGATTTTATCTGCTTCTTG 120
DB 175 GCCCGCAGCGCGGCTCGTGGTGGGATCAGCAAGTTGATTTTATCTGCTTCTTA 234

121 CATCATTTGGCACAATTTGGTCCCGAATAATTTACTTGTCTGAATGACGCCAGACTTGGAA 180
DB 235 CATCATTTGGCACAATTTAGTCCGAAAATTTATTTGCTGAGATGACGCCAGATTGGAA 294
QY 181 AAGCAGGAGAAAGTGATCAATATTTCACTCCACTCGAATGGTACTATTATTTGA 240
DB 295 AAGCAAGAGAGAGTGTACAGATGCAATCTACTCTTTGGAGTGGTACTATTATTTGA 354
QY 241 GAAGATCCAGATATTGTCTTAGAGAAATTTGAAGCAGTGGAGCAATTTCCAGCTTTGGG 300
DB 355 GAGGATCCGATATTGTCTTAGAGAAATTTAAAACACAGTGGAGCGTTCCAGTTGTGTGG 414
QY 301 AGGGTTTCAAAAGTGGAGACACACCTATTCTTTGAGGGATTTGCAATTTGATCAACA 360
DB 415 AAGGTTTTCAAAAGTGGAGAAACAATATTCCTGTAGGATTTGCAATTTGATCAACG 474
QY 361 TGTGTACTCTCTATGGACTGCTTCCAGGACAGTGTTCATAAAATCATCGTTACAGATG 420
DB 475 TGTGTCTCTGTATGGACTGCTTCCAAAGTAGTGTTCATAAAACCATCGTTACAGATG 534
QY 421 CATACTTCTACTGGAGGAGGTTCTGTGACTGTGGAGACACAGAGGCATGGAAAATGCG 480
DB 535 CATACTTCTACTGGAGGGGCTTCTGTGACTGTGGAGACACAGAGCGTGGAAAATGCG 594
QY 481 CCTTTTGTGTAAATCATCACTGGAAGAGCAGGTACTATATAAGAGAAATTCACGCTGT 540
DB 595 CCTTTTGTGTGAGTACAGAGCCTGGAAGAGCAGGTACTATCAAAAGAGAGCTTACATTGC 654
QY 541 CCGTTGAATGAAGAGGTAATTTGTCACGCCAGGAAAATATTTCCCTCAGTGTATAAATAT 600
DB 655 CCATTTGAATGAAGAGGTAATTTGTCACGCCAGGAGAAATATTCCTTCCGTGTATAAATAT 714
QY 601 GTCCTAGAAATGACTATATGGAGAGGAAAAGAACTGCCCTCCTGAATCCAGATAAGG 660
DB 715 ATTCTAGAAATGACTATATGGAGAGGAAAAGAAATTTGCCCTCCTGAACTGCAGATAAG 774
QY 661 KNR-----YVCVNDHSHYDH 675
DB 775 GAGAAAATGAACGATATATTGTCTCTCTTTTCAACAGATGAGCAACCATTCGTATGATCAT 834
QY 676 GTCATATACAGCTTACAAAGAGCTCTTCACTGTGAGCTCGCAGAGGCCAGTTCGATACC 735
DB 835 GTGATCTACAGCTTCAGAGAGCTCTAGATTCGAGAGCTTCAGAGAGCAGCTGTCACAG 894
QY 736 ACTGCCATTTGACAAAGAGGGTCTGCGGCTGTGTAAGGCGGAGCTTATCTGCTGCGCAC 795
DB 895 ACTGCCATTCGACAAAGAGGGTCTGCGGCTGTGTAAGGAGGCTGTGATGCCACTTGCAC 954
QY 796 GAACCAAGGAGATATAAGAGCTATTCAGAAAATGCTCTCAACATCCATTCATGTA 855
DB 955 GAACCAAGGAGGATATAAGAGCTATTCAGAAAATGCTCTCAAGAGAGCTTCTCAGCAAC 1014
QY 856 GAATATTACACTCAGAGATTTGCTCATCAGAAAATTTGCTTCCGCTCTGCTGCTGCTGCTG 915
DB 1015 GAATGCTGCTACTCCGCTGCTTATGCTCACCAGAAAATTTGCTTCTGCGCTTGGCTGCTG 1074
QY 916 ATGAACAAAATTTATGAGCTATTTCAGTGCATTTTAGCAGATCTTTTCCCAAGCATGCCCT 975
DB 1075 ATGAACAAAATTTATGAGCTATTTCAGTGCATTTTAGCAGATATTTTGCCAGGCTGCCCT 1134
QY 976 AGAAGAACCTGACTCGGAGAAATCCCTGTCTCATAAGCAGGTTAATGCTTTGGGATGCA 1035
DB 1135 GTAGAAGAACCTGCTCTGAAAATCCCTGTCTTATAAGCAGACTAATGCTTTGGGATGCA 1194
QY 1036 AAGCTTTATAAGGTGCCCGTAAGATCCTTCATGAATTTGATCTTCAGAGATTTTTTATG 1095
DB 1195 AAATTTTATAAGGTGCCCGTAAGATCCTTCATGAATTTGATCTTTAGTAGTTTTTTATG 1254
QY 1096 GAGATGAATACAAAACCTCTTTGCTATGAAATTTGTGAAGTATTATAAACAACCTGCAG 1155
DB 1255 GAGATGAATACAAAACCTCTTTGCTATGAAATTTGTGAAGTATTATAAACAACCTGCAG 1314
QY 1156 AAGAATATATCAGTGATGATCATGACAGAAAGTATCTCTATAACTGCATCTTTCAGTTTCAG 1215

FEATURES	Location/Qualifiers	source
BASE COUNT	969 a 546 c 694 g 850 t	
ORIGIN		
Query Match	56.2%; Score 2927.4; DB 6; Length 3059;	
Best Local Similarity	98.08; Pred. No. 0;	
Matches 2981;	Conservative 10; Mismatches 7; Indels 45; Gaps 1;	
QY	1	ATGCGGACGAGGAGCTGGAGGTACTGAGAGATGGAATCAGCGGGAGTTACCCCG 60
DB	17	ATGCGGACGAGGAGCTGGAGGTACTGAGGATGGAATCAGCGGGAGTTACCCCG 76
QY	61	ACCCTCAGCGTCTGGCATCTTGGTGGATCAGCAAGTTGATTTTATACGTCTTCTTG 120
DB	77	ACCCTCAGCGTCTGGCATCTTGGTGGATCAGCAAGTTGATTTTATACGTCTTCTTG 136
QY	121	CATCATTTGGCACAATTTGGTCCAGAAATTTACTTTGCTGAAATGGACCCAGACTTGGAA 180
DB	137	CATCATTTGGCACAATTTGGTCCAGAAATTTACTTTGCTGAAATGGACCCAGACTTGGAA 196
QY	181	AAGCAGGAGAAAGTGTACAAATGTCNAATTTCACTCCACTGGAATGGTACTTATTTGGA 240
DB	197	AAGCAGGAGAAAGTGTACAAATGTCNAATTTCACTCCACTGGAATGGTACTTATTTGGA 256
QY	241	GAAGATCCAGATATTTGCTTAGAGAAATTCAGACAGATGGAGCATTTTGCATTTGTTGG 300
DB	257	GAAGATCCAGATATTTGCTTAGAGAAATTCAGACAGATGGAGCATTTTGCATTTGTTGG 316
QY	301	AGGTTTTCAAAAGTGGAGAGACAACCTATTTTCGAGGGATTTGCAATTTGATCCAACA 360
DB	317	AGGTTTTCAAAAGTGGAGAGACAACCTATTTTCGAGGGATTTGCAATTTGATCCAACA 376
QY	361	TGTGTACTCTGTATGGACTGCTTCCAGGACAGTGTTCATATAAAATCATGTTACAAGATG 420
DB	377	TGTGTACTCTGTATGGACTGCTTCCAGGACAGTGTTCATATAAAATCATGTTACAAGATG 436
QY	421	CATACCTCTACTCGAGGAGGTCTGTGACTGTGGAGACACAGAGGCATGGAAAACCTGGC 480
DB	437	CATACCTCTACTCGAGGAGGTCTGTGACTGTGGAGACACAGAGGCATGGAAAACCTGGC 496
QY	481	CTTTTGTGTAAATCATGAACCTGGAAGACAGGTACTATATAAGAGAAATTCACGCTGT 540
DB	497	CTTTTGTGTAAATCATGAACCTGGAAGACAGGTACTATATAAGAGAAATTCACGCTGT 556
QY	541	CCGTTGAATGAAGAGTAATTTGCCAAGCCAGGAAATATTTCCCTTCAGTGATAAATAT 600
DB	557	CCGTTGAATGAAGAGTAATTTGCCAAGCCAGGAAATATTTCCCTTCAGTGATAAATAT 616
QY	601	GTGCTAGAATGACTATATGGGAAGGAGGAAAAGAACTGCCCTCGAAATCCAGATAAGG 660
DB	617	GTGCTAGAATGACTATATGGGAAGGAGGAAAAGAACTGCCCTCGAAATCCAGATAAGG 676
QY	661	KNR-----YVCVNDHSHYD 675
DB	677	GAGAAAATGAAGATACATTGTGTCTTTCAATGATGAACACCATTCATATGACCAC 736
QY	676	GTATATACAGCTACAAAGAGCTCTTGACTGTGAGCTCGCAGAGGCCAGTTGCATACC 735
DB	737	GTATATACAGCTACAAAGAGCTCTTGACTGTGAGCTCGCAGAGGCCAGTTGCATACC 796
QY	736	ACTGCCATTGACAAAAGAGGTCTGGGGCTGTAAAGCGGGAGCTTATGCTGCTTGGCCAG 795
DB	797	ACTGCCATTGACAAAAGAGGTCTGGGGCTGTAAAGCGGGAGCTTATGCTGCTTGGCCAG 856
QY	796	GAAGCAAGGAGATATAAAGAGTCATTCAGAAAAATGCTCTCAACATCCACTTTCATGTA 855
DB	857	GAAGCAAGGAGATATAAAGAGTCATTCAGAAAAATGCTCTCAACATCCACTTTCATGTA 916
QY	856	GAAGTATTACACTCAGAGATTATGCTCATCAGAAATTTGCTTCCGTCCTGGTTCCTGG 915

QY 301 AGGTTTTCAAAAGTGGAGACACACCTATTCTTCAGGGATTGTGCAATTTGATCCAAACA 360
DB |||||
QY 317 AGGTTTTCAAAAGTGGAGACACACCTATTCTTCAGGGATTGTGCAATTTGATCCAAACA 376
DB |||||
QY 361 TGTGACTCTGTATGGACTGCTTCAGACACAGTGTTCATAAAATTCATCGTTACAAAGATG 420
DB |||||
QY 377 TGTGACTCTGTATGGACTGCTTCAGACACAGTGTTCATAAAATTCATCGTTACAAAGATG 436
DB |||||
QY 421 CATACTTCTACTGGAGAGGGTCTGTGACTGTGGAGACACAGAGGCGATGAAAACCTGGC 480
DB |||||
QY 437 CATACTTCTACTGGAGAGGGTCTGTGACTGTGGAGACACAGAGGCGATGAAAACCTGGC 496
DB |||||
QY 481 CCTTTTGTGTAAATCATGAACCTGGAGAGCGGTACTATATAAGAGAAATTCACGCTGT 540
DB |||||
QY 497 CCTTTTGTGTAAATCATGAACCTGGAGAGCGGTACTATATAAGAGAAATTCACGCTGT 556
DB |||||
QY 541 CCGTTGAATGAAGAGGTAATTTGTCCAAAGCCAGGAAAATATTTCTTCAGTGATAAAATAT 600
DB |||||
QY 557 CCGTTGAATGAAGAGGTAATTTGTCCAAAGCCAGGAAAATATTTCTTCAGTGATAAAATAT 616
DB |||||
QY 601 GTCGTAGAAATGACTATATGGGAAGAGGAAAAGAACTGCTCTGAACTCCAGATAAGG 660
DB |||||
QY 617 GTCGTAGAAATGACTATATGGGAAGAGGAAAAGAACTGCTCTGAACTCCAGATAAGG 676
DB |||||
QY 661 KNR-----YVCVNDHHSDH 675
DB : : : : :
QY 677 GAGAAAATGAAAGATACTATTGTGCTCTTTCAATGATGAACACCACTTCATATGACCAC 736
DB |||||
QY 676 GTCATATACAGCTTACAAAGAGCTCTTGACTGTGAGCTCCAGAGGCCAGTTGCATACC 735
DB |||||
QY 737 GTCATATACAGCTTACAAAGAGCTCTTGACTGTGAGCTCGCAGAGGCCAGTTGCATACC 796
DB |||||
QY 736 ACTGCCATTGACAAAGAGGCTGCTGGGCTGTTAAAGCGGGAGCTTATGCTGCTGGCAG 795
DB |||||
QY 797 ACTGCCATTGACAAAGAGGCTGCTGGGCTGTTAAAGCGGGAGCTTATGCTGCTGGCAG 856
DB |||||
QY 796 GAAGCAAGGAAGATATAAGAGTCAATCAGAAAATGTCTCTCAACATCCACCTCATGTA 855
DB |||||
QY 857 GAAGCAAGGAAGATATAAGAGTCAATCAGAAAATGTCTCTCAACATCCACCTCATGTA 916
DB |||||
QY 856 GAAGTATTACACTCAGAGATATTAGGCTCATCAGAAAATTTGCTTGGCTGCTGGTCTGG 915
DB |||||
QY 917 GAAGTATTACACTCAGAGATATTAGGCTCATCAGAAAATTTGCTTGGCTGCTGGTCTGG 976
DB |||||
QY 916 ATGAACAAAATATTAGACTATTCAAGTGACTTTAGGCAGATCTTTTGGCAAGCATGCCCT 975
DB |||||
QY 977 ATGAACAAAATATTAGACTATTCAAGTGACTTTAGGCAGATCTTTTGGCAAGCATGCCCT 1036
DB |||||
QY 976 AGAAGAAGACCTGACTCGGAGAAATCCCTGCTCATAGCAGGTTAATGCTTTGGGATGCA 1035
DB |||||
QY 1037 AGAAGAAGACCTGACTCGGAGAAATCCCTGCTCATAGCAGGTTAATGCTTTGGGATGCA 1096
DB |||||
QY 1036 AAGCTTTATAAGGTGCCGTAAGATCCCTTCATGAATTCATCAGCAGTTTTCATG 1095
DB |||||
QY 1097 AAGCTTTATAAGGTGCCGTAAGATCCCTTCATGAATTCATCAGCAGTTTTCATG 1156
DB |||||
QY 1096 GAGATGGAATACAAAACTCTTTCGTATGGAATTTGTGAAGTATTATAAACAACCTGCAG 1155
DB |||||
QY 1157 GAGATGGAATACAAAACTCTTTCGTATGGAATTTGTGAAGTATTATAAACAACCTGCAG 1216
DB |||||
QY 1156 AAAGAATATACAGTGATGATCATCAGACGAAGTATCTCTATAACTGCACCTTCAGTTTCAG 1215
DB |||||
QY 1217 AAAGAATATACAGTGATGATCATCAGACGAAGTATCTCTATAACTGCACCTTCAGTTTCAG 1276
DB |||||
QY 1216 ATGTTTACTGTTCTGCTCGCATCTTATTGAAGACGAGAATGTTATCTCTGTC 1275
DB |||||
QY 1277 ATGTTTACTGTTCTGCTCGCATCTTATTGAAGACGAGAATGTTATCTCTGTC 1336
DB |||||
QY 1276 ATTTACTGAACCTCTCTAGAGTTTACTCTGAGTACTTGGACGAGAAACAATAAATTCAC 1335
DB |||||
QY 1337 ATTTACTGAACCTCTCTAGAGTTTACTCTGAGTACTTGGACAGGAACAATAAATTCAC 1396
DB |||||

QY 1336 TTCCAGGGTTATAGCCAGGACAAATTTGGGAAGAGTATATGACGTAATATCTGACCTAAAG 1395
DB |||||
QY 1397 TTCCAGGGTTATAGCCAGGACAAATTTGGGAAGAGTATATGACGTAATATCTGACCTAAAG 1456
DB |||||
QY 1396 TATATCTCTGATCAGCAAAACCCACAATATGACAGAAAGATTAAGAATGCAAGTTCCCTTGAA 1455
DB |||||
QY 1457 TATATCTCTGATCAGCAAAACCCACAATATGACAGAAAGATTAAGAATGCAAGTTCCCTTGAA 1516
DB |||||
QY 1456 GGTTCCTGATCTTTTGAAGATTTCTAGCTGTATGCAAGGAATGGAAGAAATCCGAAGA 1515
DB |||||
QY 1517 GGTTCCTGATCTTTTGAAGATTTCTACCTGTATGCAAGGAATGGAAGAAATCCGAAGA 1576
DB |||||
QY 1516 CAGTTGGCAACACATTGAACTGGATCTGATTTGGAGGCTGCCATTCGTATACAGATG 1575
DB |||||
QY 1577 CAGTTGGCAACACATTGAACTGGATCTGATTTGGAGGCTGCCATTCGTATACAGATG 1636
DB |||||
QY 1576 CAATTGAAGAATATTTTACTCATGTTTCCAAAGAGTGTGCTGTTGATCAAGAACCTCTTA 1635
DB |||||
QY 1637 CAATTGAAGAATATTTTACTCATGTTTCCAAAGAGTGTGCTGTTGATGAAGAACCTCTTA 1696
DB |||||
QY 1636 CTGTCGGCTTATAAAGAATGTCACAAAGCTGTGATGAGTGCAAGTACCAGTTTCATATCT 1695
DB |||||
QY 1697 CTGTCGGCTTATAAAGAATGTCACAAAGCTGTGATGAGTGCAAGTACCAGTTTCATATCT 1756
DB |||||
QY 1696 AGTAGCAAGACAGTAGTACAACTCCGTGTGGACATAGTTTGGAAACAAAGTCCCTACAGATG 1755
DB |||||
QY 1757 AGTAGCAAGACAGTAGTACAACTCCGTGTGGACATAGTTTGGAAACAAAGTCCCTACAGATG 1816
DB |||||
QY 1756 TCTGAGGATCTTGAAGCATACATCTGCCACTCTCTAGACCTTGTGCTGCTTTCATGTA 1815
DB |||||
QY 1817 TCTGAGGATCTTGAAGCATACATCTGCCACTCTCTAGACCTTGTGCTGCTTTCATGTA 1876
DB |||||
QY 1816 CGTTTAGCAGGCTGGGTGCTGTTTCAAGACTGCATGAATTTGCTGCTTTCATGAGACTTT 1875
DB |||||
QY 1877 CGTTTAGCAGGCTGGGTGCTGTTTCAAGACTGCATGAATTTGCTGCTTTCATGAGACTTT 1936
DB |||||
QY 1876 CAAGTAGAGTACTAGTGGAAATACCTTTACGTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1935
DB |||||
QY 1937 CAAGTAGAGTACTAGTGGAAATACCTTTACGTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1996
DB |||||
QY 1936 GCTGAGATGTGGCGAAGAAATGGACTGCTCTTATTAGCCAGGCTGCTTATTACCAAGAT 1995
DB |||||
QY 1997 GCTGAGATGTGGCGAAGAAATGGACTGCTCTTATTAGCCAGGCTGCTTATTACCAAGAT 2056
DB |||||
QY 1996 GTTAAGTGCAGAGAAAGAAATGTATGATAAGAGATATCATGCTTCAGATGGTGCATCT 2055
DB |||||
QY 2057 GTTAAGTGCAGAGAAAGAAATGTATGATAAGAGATATCATGCTTCAGATGGTGCATCT 2116
DB |||||
QY 2056 TTAATGATCCCAATAAGTTCTTGTACTGCTTACTGCTCAGAGGTATGAACCTTGCAGAGCT 2115
DB |||||
QY 2117 TTAATGATCCCAATAAGTTCTTGTACTGCTTACTGCTCAGAGGTATGAACCTTGCAGAGCT 2176
DB |||||
QY 2116 TTTAAACAAGACCATATCTACAAAAGACAGGATTTGATTAACAATATAATACACTATA 2175
DB |||||
QY 2177 TTTAAACAAGACCATATCTACAAAAGACAGGATTTGATTAACAATATAATACACTATA 2236
DB |||||
QY 2176 GAAGAAATGCTTCAGGTCCTCATCTATATTGTGGGTGAGCGTTATGTACCTGGAGTGGGA 2235
DB |||||
QY 2237 GAAGAAATGCTTCAGGTCCTCATCTATATTGTGGGTGAGCGTTATGTACCTGGAGTGGGA 2296
DB |||||
QY 2236 AATGTGACCAAGAAGAGGTCACAATGAGAGAAATCATTCACCTTGCATTTGCAATGAACCC 2295
DB |||||
QY 2297 AATGTGACCAAGAAGAGGTCACAATGAGAGAAATCATTCACCTTGCATTTGCAATGAACCC 2356
DB |||||
QY 2296 ATGCCACACAGTCCCATTTGCCAAAATTTACCTTGAGAAATGAANAATGAACCTGGCTTA 2355
DB |||||
QY 2357 ATGCCACACAGTCCCATTTGCCAAAATTTACCTTGAGAAATGAANAATGAACCTGGCTTA 2416
DB |||||
QY 2356 GAGAATGCTATAAACAAGAGTGGCCACATTTAAGAAACAGGTGTATCAGGCCATGGAGTT 2415
DB |||||
QY 2417 GAGAATGCTATAAACAAGAGTGGCCACATTTAAGAAACAGGTGTATCAGGCCATGGAGTT 2476
DB |||||
QY 2416 TATGAACATAAAGATGAATCAGTGAAGACTTCAATATGACTTTTATCATTTACTCCAAA 2475
DB |||||

Db	2477	TATGAACATAAAGATGAATCACTCAAAAGACTTCAATATGTACTTTTATCATTTACTCCAAA	2536	
Qy	2476	ACCAGCATAGCAAGGCTGAACATATGCAGAGAAAAGAGGAGAAAACAAGAAAGAT	2535	
Db	2537	ACCAGCATAGCAAGGCTGAACATATGCAGAGAAAAGAGGAGAAAACAAGAAAGAT	2596	
Qy	2536	GAAGCATTCGCCACACACACACCTCCCTGAAATCTGCGCCTCTTTCAGCAAAAGTAAAC	2595	
Db	2597	GAAGCATTCGCCACACACACCTCCCTGAAATCTGCGCCTCTTTCAGCAAAAGTAAAC	2656	
Qy	2596	CTTCTCAACTGTGATATCATGATGATACATCTCAGGACCGTATTTGAGCGGGCAATAGAC	2655	
Db	2657	CTTCTCAACTGTGATATCATGATGATACATCTCAGGACCGTATTTGAGCGGGCAATAGAC	2716	
Qy	2656	ACAGATCTTAACTTGTGGCGGAAGGATGCTCCAAATGCTTTTCATATTTGGCATG	2715	
Db	2717	ACAGATCTTAACTTGTGGCGGAAGGATGCTCCAAATGCTTTTCATATTTGGCATG	2776	
Qy	2716	GGTTTACTAGAAGAGAACACAGCTTCAAAAAGCTCCTCGAAGAAGTAACATTTGAC	2775	
Db	2777	GGTTTACTAGAAGAGAACACAGCTTCAAAAAGCTCCTCGAAGAAGTAACATTTGAC	2836	
Qy	2776	TTTTATCATAAAGCTTCAAGATTTGGGAAGTTTCAGCCATGAATATACAAATGCTTTGGAA	2835	
Db	2837	TTTTATCATAAAGCTTCAAGATTTGGGAAGTTTCAGCCATGAATATACAAATGCTTTGGAA	2896	
Qy	2836	AAACTCAAAAGGAATTCCTCCAGTTAGAGGCCAGAGGACATGATAGCTGGATCTCAG	2895	
Db	2897	AAACTCAAAAGGAATTCCTCCAGTTAGAGGCCAGAGGACATGATAGCTGGATCTCAG	2956	
Qy	2896	ATGTTTGACACAGTGAAGCGATTAAAGAGAAAAATCTTTGTTAAATTTGTAGCAACACATCA	2955	
Db	2957	ATGTTTGACACAGTGAAGCGATTAAAGAGAAAAATCTTTGTTAAATTTGTAGCAACACATCA	3016	
Qy	2956	GGATCGGAATCTATTAGAATGATGAGATTACTCATGATTAAG	2998	
Db	3017	GGATCGGAATCTATTAGAATGATGAGATTACTCATGATTAAG	3059	
RESULT 8				
BD156870		2550 bp	DNA	linear
LOCUS	BD156870			PAT 17-JAN-2003
DEFINITION				Primer for synthesizing full-length cDNA and use thereof.
ACCESSION	BD156870			
VERSION	BD156870.1	GI:27862628		
KEYWORDS	JP 2002191363-A/11713.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE				
AUTHORS	Ota,T., Isoqai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J., Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.			
TITLE	Primer for synthesizing full-length cDNA and use thereof			
JOURNAL	Patent: JP 2002191363-A 11713 09-JUL-2002;			
COMMENT	HELIX RESEARCH INSTITUTE			
	OS Homo sapiens (human)			
	PN JP 2002191363-A/11713			
	PD 09-JUL-2002			
	PF 28-JUL-2000 JP 2000280990			
	PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU SAITO,			
	PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,			
	PI KEIICHI NAGAI, TETSUO OTSUKI			
	PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/10,			
	PC C12P1/02, C12N1/68/ C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC			
	Primer for synthesizing full-length cDNA and use thereof			
	PH key			
	Location/Qualifiers			
FT	CDS	(117)..(2549).		
FEATURES				
	Location/Qualifiers			


```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="PLACE1004743"
/tissue_type="placenta"
/clone_lib="PLACE1"
/note="cloning vector: pME18FL3"
117..>2550
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB55380.1"
/db_xref="GI:14042752"
/translation="MTWEEGRKVPPELQIREK
ALDCEALQALHTTATIDGEGRAKVAAGAAYCQ
ETMAHQFALRLGSGWANKIMSYSDRQIFCOAQ
KGARLKLHIFSSQFMWEMYSKKLFAMEYKYY
FVPTLARHLTEQNVASVTETLLEVLPEYLD
KYLTLKPTITERLQWQLEGFPSFKLITCQM
IOMQLKTLMLPQWCAQDELLVLAKECHA
KSTRVSEDVLSIHLPSTLAGLHVRLSRGAV
VLAQVVAQWMRRNGLSLSQSVFYQDVKREK
LQRYELAEAFNTSTKDQDLKOYNTLLEML
RETILHLCIEDPMSHATSNENNETNGLEN
DLDFNNYFVHKYKTOHSHKAEHQKRRQENH
IMMYLITVFERADITDSNLTWGMLOKAFKIL
KASRLGSSAMINQMLLEKLAGLPQLEGQDMITT
SESLKQDITTHDK"

```

Query Match	46.8%;	Score 2434.4;	DB 9;	Length 2550;
Best Local Similarity	97.6%;	Pred. No. 0;		

QY	494	ATCATGAACCTGGAAGACGAGGTACTATAAAAGAGAAATTCACGCTGCTCCGTTGAATGAAG	553
Db	1	ATCATGAACCTGGAAGACGAGGTACTATAAAAGAGAAATTCACGCTGCTCCGTTGAATGAAG	60
QY	554	AGGTAATTGTCACGAGCCAGGAAATATTTCCCTTCAGTGATAAAAATATGTGCTAGAAATGA	613
Db	61	AGGTAATTGTCACGAGCCAGGAAATATTTCCCTTCAGTGATAAAAATATGTGCTAGAAATGA	120
QY	614	CTATATGGGAGAGGAAAAAGAACTGCTCTCGAACCTCCAGATAAAGKNR	663
Db	121	CTATATGGGAGAGGAAAAAGAACTGCTCTCGAACCTCCAGATAAAGGAGAAAAATGAAA	180
QY	564	-----YYCVDNHSYDHGTCATATACAGCC	688
Db	181	GATACTATTGTGTCTTTTCAATGATGAACACCATTCATATGACACGCTCATATACAGCC	240
QY	689	TACAAAGAGCTCTTGACTGTGAGCTCCAGAGGCCAGTTGCATPACCACATGCCATTGACA	748
Db	241	TACAAAGAGCTCTTGACTGTGAGCTCCAGAGGCCAGTTGCATPACCACATGCCATTGACA	300
QY	749	AAGAGGGTCCTCGGGCTGTTTAAAGCGGGAGCTTATGCTGCTTGCCAGGAAGCAAAAGGAAG	808
Db	301	AAGAGGGTCCTCGGGCTGTTTAAAGCGGGAGCTTATGCTGCTTGCCAGGAAGCAAAAGGAAG	360
QY	809	ATATAAGAGTCATTGAGAAAATGTCTCAACATCCACTTCATGTAGAGTATTACACT	868
Db	361	ATATAAGAGTCATTGAGAAAATGTCTCAACATCCACTTCATGTAGAGTATTACACT	420
QY	869	CAGAGATTATGGCTCATCAGAAAATTTCCCTTTGGGCTTGGTTCTCGATGAACAAAAATTA	928
Db	421	CAGAGATTATGGCTCATCAGAAAATTTCCCTTTGGGCTTGGTTCTCGATGAACAAAAATTA	480
QY	929	TGAGCTATTCAAGTGACTTTTAGGCAGATCTTTTGCCAAAGCATGCTTTAGAGAAGAACTG	988
Db	481	TGAGCTATTCAAGTGACTTTTAGGCAGATCTTTTGCCAAAGCATGCTTTAGAGAAGAACTG	540
QY	989	ACTCGGAGAAATCCCTGTCTCATTAAGCAGGTTAATGCTTTGGGATGCAAAAGCTTTTATAAG	1048
Db	541	ACTCGGAGAAATCCCTGTCTCATTAAGCAGGTTAATGCTTTGGGATGCAAAAGCTTTTATAAG	600

Qy	1049	GTGCCGTAGATCCCTTCATGAATTTGATCTTCAGCAGCTTTTTTTTATCGAGATGGAATACA	1108
Db	601	GTGCCGTAGATCCCTTCATGAATTTGATCTTCAGCAGCTTTTTTTTATCGAGATGGAATACA	660
Qy	1109	AAAAACCTTTTGTCTATGGAATTTGTGAAGTATATATAAACAACTGCACAAAGAATATATACA	1168
Db	661	AAAAACCTTTTGTCTATGGAATTTGTGAAGTATATATAAACAACTGCACAAAGAATATATACA	720
Qy	1169	GTGATGATCATGACAGAAGTATCTCTATAACTGCACCTTCAGTTTCAGATGTTTACTGTTTC	1228
Db	721	GTGATGATCATGACAGAAGTATCTCTATAACTGCACCTTCAGTTTCAGATGTTTACTGTTTC	780
Qy	1229	CTACTCTGGCTCGACATCTTATTGAAGACAGAAATGTTATCTCTGTGTCATTACTCGAAACTC	1288
Db	781	CTACTCTGGCTCGACATCTTATTGAAGACAGAAATGTTATCTCTGTGTCATTACTCGAAACTC	840
Qy	1289	TGCTAGAAGTTTACCTGACTACTGTGCAGAGGAACAATAAATTCACCTCCACGGTTTATA	1348
Db	841	TGCTAGAAGTTTACCTGACTACTGTGCAGAGGAACAATAAATTCACCTCCACGGTTTATA	900
Qy	1349	GCCAGGACAAATTTGGGAAGAGTATATGCAGTAAATATGTGACCTAAAGTATATCTCTGATCA	1408
Db	901	GCCAGGACAAATTTGGGAAGAGTATATGCAGTAAATATGTGACCTAAAGTATATCTCTGATCA	960
Qy	1409	GCAACCCCAATATGGACAGAAAGATTAGAATGCGAGTTCCTTTGAAGGTTTTCGATCTTT	1468
Db	961	GCAACCCCAATATGGACAGAAAGATTAGAATGCGAGTTCCTTTGAAGGTTTTCGATCTTT	1020
Qy	1469	TTTTTGAAGATTCCTACCTGTATGCAGGGAATGGAAGAAATCCGGAAGACAGGTTGGGCAAC	1528
Db	1021	TTTTTGAAGATTCCTACCTGTATGCAGGGAATGGAAGAAATCCGGAAGACAGGTTGGGCAAC	1080
Qy	1529	ACATTGAAGTGGATCTCTGATTGGGAGGCTGCCATTGCTATACAGATGCAATTGGAAGAATA	1588
Db	1081	ACATTGAAGTGGATCTCTGATTGGGAGGCTGCCATTGCTATACAGATGCAATTGGAAGAATA	1140
Qy	1589	TTTTTACTCATGTTTCCAAAGAGTGTGTGCTTTGATGAAGAACTCTTACTTTGTGGCTTTATA	1648
Db	1141	TTTTTACTCATGTTTCCAAAGAGTGTGTGCTTTGATGAAGAACTCTTACTTTGTGGCTTTATA	1200
Qy	1649	AAGAAATGCACAAGCTGTGATGAGTGCAGTACCAGTTTCATATCTAGTAGCAAGACAG	1708
Db	1201	AAGAAATGCACAAGCTGTGATGAGTGCAGTACCAGTTTCATATCTAGTAGCAAGACAG	1260
Qy	1709	TAGTACAATCGTGTGGACATAGTTTGGAAACAAGTCCCTACAGAGTATCTCAGGATCTTG	1768
Db	1261	TAGTACAATCGTGTGGACATAGTTTGGAAACAAGTCCCTACAGAGTATCTCAGGATCTTG	1320
Qy	1769	TAGACATACATCTGCCACTCTCTAGGACCCCTGCTGGCTCTTCATGTACGTTTAAAGCAGGC	1828
Db	1321	TAGACATACATCTGCCACTCTCTAGGACCCCTGCTGGCTCTTCATGTACGTTTAAAGCAGGC	1380
Qy	1829	TGGGTGCTGTTTCAAGACTGCATCAATTTCTGTCTTTTGGAGCTTTTCAAGTAGAGGTAC	1888
Db	1381	TGGGTGCTGTTTCAAGACTGCATCAATTTCTGTCTTTTGGAGCTTTTCAAGTAGAGGTAC	1440
Qy	1889	TAGTGGAAATACCTTTACGTTTGCTGGTGTGGTTGGCCAGGTTGTTGCTGAGATGTGGC	1948
Db	1441	TAGTGGAAATACCTTTACGTTTGCTGGTGTGGTTGGCCAGGTTGTTGCTGAGATGTGGC	1500
Qy	1949	GAAGAAATGGAGTCTCTCTTTATTAGCCAGGTGTTTTATTACCAAGATGTTTAAAGTGCAGAG	2008
Db	1501	GAAGAAATGGAGTCTCTCTTTATTAGCCAGGTGTTTTATTACCAAGATGTTTAAAGTGCAGAG	1560
Qy	2009	ACAAGATGTATGATAAAGATATCATCATGCTTCAGATTTGGTGCATCTTTTAAATGGATCCCA	2068
Db	1561	ACAAGATGTATGATAAAGATATCATCATGCTTCAGATTTGGTGCATCTTTTAAATGGATCCCA	1620
Qy	2069	ATAAGTCTCTTTTACTGGTACTCTCAGAGGTATGAACCTTGCCGAGGCTTTTAAACAAGACCA	2128
Db	1621	ATAAGTCTCTTTTACTGGTACTCTCAGAGGTATGAACCTTGCCGAGGCTTTTAAACAAGACCA	1680
Qy	2129	TATCTACAAAAGACCAGGATTTGATTTAAACAATATAATACACTAATAGAAGAAATGCTTC	2188

Db 322 CCTACATATCTTGCAGAGACTGTGCAGTGTGATCCAACTTGTTGTCATGAGTGC 381
Qy 382 TTCCAGGACAGTGTTCATAAATCATCGTTACAGATGCATCTTACTTGGAGGG 441
Db 382 TTTTGGGAATATTCACAGAGATCATCGATAGGATGACAACATCAGGAGGTGGAGT 441
Qy 442 TTCTGTGACTGTGGAGACACAGAGCATGGAAACTGGCCCTTTTGTGTAATCATGAA 501
Db 442 TTCTGTGACTGTGGATGACTGAAGCCTGGAAAGGGTCTTACTGTCAAAACATGAA 501
Qy 502 CTTGGAAGAGAGAGTACT--ATAAAGAGAAATTCACGCTCTGCGTTGAAATGAAGAGTA 558
Db 502 CTTAACACCTCTGAATTTGAGGAAGAGAGGATCCTCTGTTCATTTATCAGAAGATGTG 561
Qy 559 ATTGTCCAGCCAGGAAATATTTCCCTTCAGTGATAAATATGTCGTAGAAATGACTATA 618
Db 562 ATAGCAAGAATTTATACATTTTGTCTATTACGTTTCGGTATGTCAGTGAATAATTAACC 621
Qy 619 TGGGAAGAGGAAAGAACTGCCCTCCCTGAACCTCCAGATAAGKNRY----- 664
Db 622 TGGGAAGAGAAAGTGAATGCCAGCAGATTTAGAGATGGTAGAGAGAGTGACACCTAC 681
Qy 665 -----YCVNDHSHSYDHGTCATACAGCCCTACAAA 693
Db 682 TATTGCATGCTGTTTAAATGATGAGTTTCACACCTATGAACAAGTTATTTATCTTTCAG 741
Qy 694 AGAGCTCTTGACTGTGAGCTCGCAGAGCCAGTTGTCATACACTGCCATTCACAAAGAG 753
Db 742 AAAGCTGTAACTTACACAAAGAAAGCTATTGGTTTGGAACTTACAGTAGATCGAGAT 801
Qy 754 GGTGCTGGGCTGTTAAAGCGGAGCTTATGCTGCTGCCAGGAGCAAGAGATATA 813
Db 802 GGGCGTAGTCTGTGATGAGATTTTCAGTATTGTGACGACGAAATCAGTAAT 861
Qy 814 AAGAGTCATTACAGAAATGTCCTCAACATCCACTTCATAGAGATTTACACTCAGAG 873
Db 862 GTGAGAAATACCACTAGACAGAC--AAAGCCACTCAAGTTCAGTTATGCATTCGTCT 918
Qy 874 ATTAGGCTCATCAGAAATTTGCTTTGGTCTTGGTCTCGATGAGCAAAATATAGAC 933
Db 919 ATTGTCACATCAGAAATTTGGTTTGAACATTTTGTCTGGCTGGGAAGTATTATTGGA 978
Qy 934 TATTCAAGTGACTTTAGGCAGATCTTTTGCCAGCATGCCCTTAGAGAGAACCTGCAGCG 993
Db 979 TATTCAAGTGCCCTTCGCCGATTTTATGTCAGTTGGTTTACAGAAAGGCCAGATGGT 1038
Qy 994 GAGATCCCTGTCTCATAAGCAGGTAAATGCTTTGGGATGCAAGCTTTATAAGGTGCC 1053
Db 1039 GAAACTCTCTCTAGTGGACAGACTGATGCTTAGTGATCCAAATTTAGGAAAGGTGCT 1098
Qy 1054 CGTAAGATCCCTTCATGAATGATCTTCAGCAGTTTTTTTATGGAGATGGAATACAAAAA 1113
Db 1099 AGGAGTCTATATCATCAGTTGTTTCATGACAGCTGCTTATGGATTTGAATATCAGAA 1158
Qy 1114 CTCTTTGCTATGGAATTTGGAAGTATTATAAACAACCTGCAGAAAGAAATATCATGTGAT 1173
Db 1159 CTATTTGCTGTGATTTGCAAAAAATTTACCAGCAGTTGCAGAGAGATTTTATGAGGAT 1218
Qy 1174 GATCATCAGAGAAATCTCTATAACTGCACCTTTTCAGTTTCAGATGTTTACTGTTCCCTACT 1233
Db 1219 GATCAGAGGCGAGAGTGTGCGGTGACTGCTATCTATCTGCCAGTTCTTCACCGCACCTACT 1278
Qy 1234 CTGGCTCGACATCTTATTGAAGAGCAGAACTGTTATCTGTCATTACTGAAACTCTGCTA 1293
Db 1279 CTGGCTCGAATGCTCATCAGAGAAACCTTGATGAGCAATATCAATTAAGACTTTTATG 1338
Qy 1294 GAAGTTTT-----ACCTGAGTACTTGGACAGCAACATAAATTCACCTCCAGGGTTAT 1347
Db 1339 GATCATTTGACACATCGAGTGCACGAGAGATTTTCAAGTATGACGATACACTGCTTTA 1398
Qy 1348 AGCCAGGACAAATTTGGAAGAGTATATGCACTAATATGTGACCTAAAGCTATATCTGATC 1407
Db 1399 CAAGCCTTCAAAATTTAGGAGAGTACAGAGCCTTATTTTATGATCTCAAGTATGTTTAAT 1458

Qy 1408 AGCAABCCACAAATATGAGCAGAAAGATTAAGAAATGCAGTTCCTTGAAGCTTTTCGATCT 1467
Db 1459 AGCAABCCAACTGAATGGTCAGATGAGCTGAGCAGAAAGTTCTCTAGAAAGGTTTGAAGCC 1518
Qy 1468 TTTTGAAGATCTTACCTGTATGCAGGGAATGGAAGAAATCCGAAGACAGGTGTGGCAA 1527
Db 1519 TTTTGAATTAATAAAATGATGAGGGAATGATCCCAATACACCTCAAGTAGGACAA 1578
Qy 1528 CACATTGAAGTGGATCCTGATGGAGGCTGCCATGCTATACAGATGCAATTAAGAAT 1587
Db 1579 CATATTGAATGGAACAGAGTGGAAAGCAGCTTCACACTACAAATGAATTAACACAT 1638
Qy 1588 ATTTTACTCATGTTTCAAGAGTGGTGTGCTTGTGATGAAGAACTCTTACTTGTGCGCTTAT 1647
Db 1639 GTCAATTTCAATGATCAGGAGCTGGTGTGCTTCAGATGAAAAGTGTTAATCGAAGCTTAC 1698
Qy 1648 AAAGAAATGTCACAAAGCTGTGATGAGTGCAGTACCAGTTTTCATATCTAGTAGCAAGACA 1707
Db 1699 AAGAATGTCCTGCTGCTACTGATGAGTGCAGTGTGCTGTTATCTGATGAGTGAACAGCCA 1758
Qy 1708 GT---AGTACAAATCGTGTGACATAGTTTTGGAAACAAAGTCTTACAGAGTATCTGAGGAT 1764
Db 1759 ATCACAATAAGCATTTTGTGACATTCAGTGAACATATCAGATACTGTGTTTCCCAAGAA 1818
Qy 1765 CTTGTAAGCATACATCGCACTCTCTAGAACCTTGTGCTGCTTCTATGATGAGTGTAAAGC 1824
Db 1819 AAAGTTAGCATTCACCTCCAGTTTCTCGCTTACTTGCAGGTTTACATGATATTATTAAAGC 1878
Qy 1825 AGGCTGGGTGCTGTTTCAAGACTGCATGAATTTGTGCTTTTGGAGGACTTTCAAGTAGAG 1884
Db 1879 AAAAGTGAAGTGCATATAAATTTCCAGAGCTCCTACCTCTAAGTGAACCTTAGCCCAACC 1938
Qy 1885 GTACTAGTGGAAATATCCTTTTACGTTGTCTGCTGTTGGTGTGCCAGGTTGTGCTGAGATG 1944
Db 1939 ATGTTGATAGAAACACCTCTTAGATGCTTGTGCTGTGTCGCCAAGTACATGCCGAATG 1998
Qy 1945 TGGGGAAGAAATGGAGTCTCTTATAGCCAGGTGTTTATTACCAAGATGTTAAAGTGC 2004
Db 1999 TGGGAAGAAATGGGTCTCTCTAGTAAACAGATTTATTACTACCAATAATGTGAAATGC 2058
Qy 2005 ACAGAAAGAAATGATGATAAGATATCATCATCTTTCAGATGGTGATCTTTAATGAT 2064
Db 2059 AGACGTGAGATGTTTGACAAGGATGATGTAATGCTTTCAGACAGGTGCTCCATGATGAT 2118
Qy 2065 CCAATAAGTCTCTGTTTACTGTTGTTTACAGAGTATCAAGTTCGAGGCTTTTAAACAAG 2124
Db 2119 CCAATCATTTCTGATGATCATGCTCAGCCGCTTTGAACTTTATCAGATTTTTCAGTACT 2178
Qy 2125 ACCATATCTTACAAA-----GACCAGGATTTTGAATTAACAATAT 2163
Db 2179 CCAGACTATGGAAGAAAGATTTAGTCTGAGATTAACCATAAAGGATGTTTTCAGCAGAAC 2238
Qy 2164 AATACACTAATAGAGAAATGCTTCAGGCTCCTCATCTATATATTGTTGGGTGAGCGTTATGTA 2223
Db 2239 AATACTCTAATAGAGAAATGCTTATACCTCATTAATGCTTGTGAGAGAGATTTAGT 2298
Qy 2224 CCTGAGTGGAAATGTGACCAAGAGAGGTGCACATGAGAGAAATCATCTACTTCTGCT 2283
Db 2299 CCTGAGTGGACAGGTAATGCTACAGATGAATCAAGCGAGAGATTTATCCATCAGTTG 2358
Qy 2284 TCATTGAACCCATGCCACACAGTGCATTTGCCAAAAATTTTACTGAGAATGAAAATAAT 2343
Db 2359 AGTATCAAGCCTATGGCTCATAGTGAATTTGGTAAAGTCTTTACTGAGATGAGAACAG 2418
Qy 2344 GAACTGGCTTAGAGAAATGTCATAAACAAGTGGCCACATTTAAGAAACACAGGTGATCA 2403
Db 2419 GAGACTGGCATGGAGAGTGTAACTGAAGCAGTTGCCCATTTCAAGAAACCTGGATTAA 2478
Qy 2404 GGCATGGAGTTTGAACATAAAGATGAATGATCACTGAAAGACTTCAATATGFACTTTTAT 2463
Db 2479 GGACGAGGCGATGTATGAACCTGAAACCAAGATGTCACCAAGAGTTCACTTGTATTTCTAT 2538

Qy	2464	CATTACTCCAAAACCCAGCATAGCAAGCGCTGAACATATGCGAAGAAAGAAAAGAGAAAACAA	2521
Db	2539	CACTTTTTCCAGGGCAGACAGTCTCCCAAGGCGAAGAACGAAATTTGAAAGACAA	2598
Qy	2524	GAAAACAAAGATGAAGCATTCGCCGCCACACCACCTCCTGAATTCCTGCCCTGCTTTCAGC	2583
Db	2599	AATAGAAAGATACAGCAGCTCCCAACCTCCCGGTGTGCTCCATCTCTGCCCTCTGTTGCA	2658
Qy	2584	AAAGTGATTAACTTCTCAACTGTGTATATCATATGATGTACATCTCTCAGGACCGTATTGAG	2643
Db	2659	AGCCTGGTTAACATTTTTGCACTCAGATGTCTATGTTGTGCATCATGGGAACAATTCGCAA	2718
Qy	2644	CGGCAATAGACACAGATTCTAACTTGTGGACCGAAGGGATGCTCCAAATGGCTTTTCAT	2703
Db	2719	TGGCTCTGGAAACATAATGGATATGCTTGCTCAGAGTCCATGCTGCAAAAGGGTGTACAT	2778
Qy	2704	ATTCTGCGATTGGGTTTTACTTAGAAGAGAACACAGCTTCAAAGAGCTCTCAAGAGCTCTGAAG	2760
Db	2779	TTAATTGGCATGGCAGCTACAGAAGAAAACAACATTTAGAGAATGTCAACGGAAGAGCAT	2838
Qy	2761	GAAGTAACATTTGACTTTTATTAAGAGCTTCAAGATTTGGGAAGTTTCAGCCATGAAT	2817
Db	2839	GTAGTAACATTTACCTTCACTCAGAGATATCAAAACCTGGTGAAGCCCAAAAATCT	2898
Qy	2818	-----ATACAATATGCTTTTGGAAAACTCAAAGGAATCCCCAGTTTAGAAGCCAGAAG	2871
Db	2899	CCTAGCATATAGTATGCTGGAAACACTACAAAATGCTCCCTACCTAGAGTCCACAA	2958
Qy	2872	GACATGATAACGTGGATACTTCAGATGTTTGACACAGTGAAGCGATTAGAGAAAATCT	2931
Db	2959	GACATGATCGGTGGATATTGAAGACTTTTAAATGCTGTTAAAAGATGAGGGAGAGTTCA	3018
Qy	2932	TGTTTAATTGTAGCAACCACATCAGGATCGGAATCTATTAAAGATGATGAGATTACTCAT	2991
Db	3019	CCTACCACTC-----CCGTGGCAGACAGAGAAGAACCAATAATGAAGAGAGTTCAAGG	3072
Qy	2992	GATAAGAAAAGCAGAACGAAAAGAAAAGCTGAAGCTGCTAGGCTACATCGCCAGAAG	3051
Db	3073	GACAAAGACAAAGCTGAGAGGAAGAGAAAAGCAGAGATTGCCAGACTGGCCACAGAAAG	3132
Qy	3052	ATCATGCTCAGATGTCTGCTTTACAGAAAACCTTCATTGCAAACTCATAAACTCATGTAT	3111
Db	3133	ATCATGCTCAGATGTCTGAAATTCGACGGCGATTTTATTGATGAAAACAAGAAACTCTTT	3192
Qy	3112	GACAATACATCAGAAATGCCTGGGAAGAAGATTCCATTATGAGGAAGAGACGCCCA	3171
Db	3193	CAGCAGACATTAGAACTGGATGCCTCAACCTCTGCTGTTCTTGATCA-----TAGCCCT	3246
Qy	3172	GCAGTCACTACTCTAGAAATGCTTTGGGTCTCTAAAGGGGTCCATCTGTACTGAA	3231
Db	3247	GTGGCTTCAGATATGACACTTTACAGCACTGGGCCCGCACAAACTCAGGTTCTCGAACAA	3306
Qy	3232	AAGGAGTGTCTGACGTGCATCCTTTGCCAAGAGAACAGGAGGTGAATAAGAAAATAT	3291
Db	3307	AGCAATTCGTTACATGTATTGTGTCAAGAGAGCAAGAAGTTAAAGTGGAAAGCAGG	3366
Qy	3292	GCCATGTTATTATCGGCTGTGCCAGAAATCTACTGCCTTAAACCAGACAGCGGGA	3351
Db	3367	GCAATGTCTTGGCAGCATTTGTTTCAGAGATCAACTGTATTATCAAAAACAGAAAGTAA	3426
Qy	3352	CCCATAGAACTCTCAGAGAAAGCCCTTAGACCCACTTTTTCATGTGATCCAGACTTGGCATAT	3411
Db	3427	TTTATTCAAGATCCAGAAAAA---TATGATCCATTATTCATGCACCTGATCTGTCTGT	3483
Qy	3412	GGAACTTATACAGGAAGCTGTGTGTCATGTAAATGCCAGCAGTGTGCTGGCAGAAGTATTTT	3471
Db	3484	GSAAACACACTAGTAGCTGGGCACATTTATGCATGCCCATTTGTTGGCAAGGTTATTT	3543
Qy	3472	GAAGCTGTACAGCTGAGCTCTCAGCAGCGCATTCATGTTGACCTTTT-----T	3519
Db	3544	GATTCGGTTCAAGCTAAAGAACGGAAGCAAGAGATATACGCTTTACATACGAGCTAT	3603
Qy	3520	GACTTGGAAAGTGGAAATATCTTTGCCCTCTTTGGCAATCTCTGTGCAATACTGTGATC	3579

[illegible]

```
Db 4660 AACAACTCATTTGGCTTTTCAAGAAAATAGTGAGATATGAATTCACGTGATGAAAGT 4719
QY 4657 CGGTGTGAGATCTCCCTACTTAAACTGTGTTGAAGCAAAAACACCCTGGTCAGGTAC 4716
Db 4720 TGGTCCGTAACAGTGAAGTAAAGATATAGAGAGTCAAGAGATGCTATAAGATAT 4779
QY 4717 CCTAGAAAAGAAATAGTTGTATAGAGCTTCTGTATAGCTATATAGCTGCTCCTGTAATCAA 4776
Db 4780 CCAAGAGAACTAACAATAATTAATAACCTTCCAGAGAGATATACAGAGCCTCATTAATCAA 4839
QY 4777 GCTTCTCATTTCCAGGTGCCACGGTCTGCAGATGATGAGGAAGCATCCTGTCCTCTGC 4836
Db 4840 GCATCAATTTCTGTCGCCGGAATCAGGTGGTGATAGAGCAGACCCCAACTCTGTGC 4899
QY 4837 CTTTCTGTGGGCTATATCTGTTCTCAGAACATTTGCTGCCAGAAATTTGTGAACGGG 4896
Db 4900 CTTGTGTGGGATCTCTGCTGCTCCAGAGTTACTGCTGCCAGACTGAACTGGAAGGG 4959
QY 4897 GAAGAGTTGGAGCTGCAATTTTTCAGGCACTTCTGAGGAGCCGGAGTCTGCATTTTC 4956
Db 4960 GAGGATGTAGGAGCCTGCACAGCTCACACTACTCTGTGGCTCTGGAGTGGCATCTTC 5019
QY 4957 CTAATAATCAGAGATCGGAGTGGTCTGTTGAGGTAAAGCCAGAGGCTGTCCTAT 5016
Db 5020 CTGAGAGTACGGGAATGTCAAGTGCTATTTTAGTGGCAAAACCAAGGCTGTTTTAT 5079
QY 5017 CCAGCTCTCTACTTGGATGAATATGAGAAACAGACCTTGGCTGAGAGGGCAACCCC 5076
Db 5080 TCTCTCTTACTTGTGACTATGAGGAGACCGACAGGACTCAGCGGGGAATCTCT 5139
QY 5077 CTTCAATTTATCTGTCGAGCGGATCGGAAGCTCCATTTGCTGCGCAACAACTGCATT 5136
Db 5140 TTACATTTATGCAAGAGCGATTCAGAGATTCAGAGCTCTGGCACCAACACAGTGT 5199
QY 5137 ATAGAGAGATTGCTAGGAGCAAGAGACTAATCAGATGTTATTTGGATTCAACTGGCA 5195
Db 5200 ACAGAGGAATTTGGACATGCACAGGAAGCAATCAGACACTGTTGGCATTTGACTGCA 5258
```

RESULT 11

```
AY061885
LOCUS Mus musculus ubiquitin ligase E3 alpha-II mRNA, partial cds.
DEFINITION Mus musculus ubiquitin ligase E3 alpha-II mRNA, partial cds.
ACCESSION AY061885
VERSION AY061885.1 GI:27434481
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 5265)
AUTHORS Han,Yotya; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Novel ubiquitin ligase E3 alpha-II
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 5265)
AUTHORS Han,H.Q. and Kwak,K.
TITLE Direct Submission
JOURNAL Submitted (05-NOV-2001) Research, Amgen Inc., One Amgen Center
Drive, Thousand Oaks, CA 91320, USA
FEATURES
source
1..5265
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C3H/HeN"
/db_xref="taxon:10090"
/dev_stage="4-6 weeks old"
1..5265
/codon_start=1
/product="ubiquitin ligase E3 alpha-II"
/protein_id="AAL32102.1"
/db_xref="GI:27434482"
/translation="MASEMEPEVQALDRSLLECSABEIAGRWLAQTLDRNRYVQHLAH
CVPKIYCRGNPFPQKEDTIAQHLILGPMEWYICAEADPALGPKLEQANKPSHLCGRV
```

CDS

```
FKVGEPTYSCRDCADVPTCVLCMECFILGSHRDHRYRMTTSGGGGFCDCGDTAEAWKEG
PYCKHKLSSVEVVEEDPLVHSEDIYATNTNIPAFIRYAVDILVTEKSESEPLEDL
EVAEKSTYYTCLMFNDSEVHTYEQVYITLQRAVNTQKFAIGFATVDRDGRPPRYGD
FOYCDQAKTVIVRNTSOTKPLKVOVMHSSVAAHONFGLKALSGLWSVIGYSDGLRRI
LCQVGLQEGPDGNSLVDRLMLNDSKLWKGARSYVHOLPMSSLLMDLYKKLPAALRF
AKNYROLQROPFMEDDHRAVSVALSVQFTTAPTLARMLLITEENLMTVILKAFMDHLK
HRDAQGFQFERYTALQAFRRVQSILDLKLYVLSKPTIEMSDRELROKFLQGGDAFL
ELKLCMGMDPIITRQVGHIMEPEEAAFTLQMKLTHVISMDWCADKDEKVLIEAY
KSKVAYTQCHGGTDEGPITLSICGHSVETIRYCVSEKVSILHPSRLLAGLHLV
LKSLEAVYKPELLPLSELSPULIEHPLRCLVLCVAOVHAGMRNRNGSLNQVLYYH
NVKCRMEFPKDIVMLQTVGSMMDPNHFLMILSRFELYOLFSTPDYKGRSSEVTHK
DVYQONTLLEMLYLLIMLVGERFPGVGVAAATDEKREILHQLSLKPMASELVK
SLPEDENKETGMSVIESVAHFAPKPTLGRMYELKPECAKEFNLYTFHFRAGQSLK
EWAQRKLRENKEDTALPPALPPCFPLFASLVNLOCDVMLYIMGTTLQWAVEHHGS
AWESMLQRYLHLITGMALEKHEHLENAVEHVQTFQTKISKPDAHPNSPILAM
LETLQNAPLSEAKHMDMLWLNKFNKIKRECSSQSPVAEAGT IMESRRDKDAE
RKKKAEIARLRREKIMAOSEMRHFIDENKELFOOTLEDTLSATLSDSPVSDAA
LRTALGPAQTVQPEPRQVTCILCOEEQEVTVSRAMVLAFAVQSTVLSKQRTKIAD
PEKIDPLFMPDLSGTHGSCGHVMAHCWQRYFDSVQAKRQRQRLRLTSDVE
NGBFLCPLCELSNTVITPLLPRLSILSRNLNFSQDPLAQMTRAVTOOIKVQWMLR
KHAADTSSSEDTEAMNLIPIPEGRFPDFPNPYSDSIKEMLTFTGTAAYKVLKVH
PNEADPRVPILCWGTCATYTIQSIERILSDEKVPGLPCLRDLCLSLTRFAAAHWT
VALLPVVQGHFCKLFASIVPSDSYEDLPCILIDIMFHLVGLVLAFLALQCDQFSSSS
LATGDHLIFHLVTHAHIVQILLTSCTEENGMDQNPTEGEBELAILSLKTHQYTGSA
LKBAQSWHLWRSVRAAAMPKCSALFFHYLNGVPAPPDLOVQSGTSHFHLNCLLSL
INOASNFCPSKGGDKRAPTLCLVCSGLCSQYCOALEGDEVGACTAHTYSCGS
GAGILFVRECOVLFLAKTKGCFYSPYLDYDGETDGLRRGNPLHLQCERFRKIQK
LWQHSITEIGHQAEANQILVGDWQHL"
BASE COUNT 1402 a 1269 c 1369 g 1225 t
ORIGIN
Query Match 22.18; Score 1151.8; DB 10; Length 5265;
Best Local Similarity 54.08; Pred. No. 4.7e-270;
Matches 2825; Conservative 9; Mismatches 2247; Indels 150; Gaps 18;
QY 74 TGGCATCTTGGTGGGATCAGCAAGTTGATTTTTTATCTGCTTTCTTGCATCATTTGGGAC 133
Db 71 TCGCAGGGAGATGCTGCACCAACCGACCTCAACAGAGAAGTGTACCAGCATTTAGCCC 130
QY 134 AATTGGTCCAGAAATTTACTTTTCTGTAATGGACCCAGACTTGGAAAGCAGGAGGAAA 193
Db 131 ACTGTGTGCCAAAATCTACTGCGGGGCCCTAAACCCCTTCCCTCAGAAGGAAGACACGC 190
QY 194 GTGTACAAATGTCAATATTCACTCCACTGGAAATGGTACTTATTTGGAGAAAGTCCAGATA 253
Db 191 TGGCACAGACATCTCTGCTGGACCGGATGGTACTATCTGCGCTGAAGACCTGGCG 250
QY 254 TTTGCTTAGAGAAATTTGAAGCAC---AGTGGAGCATTTTCAGCTTTGTGGAGGGTTTCA 310
Db 251 TGGGATTTCAAAGCTCGAGCAGGCAACACAGCCCTTCACTCTGTGGCCGAGTGTTA 310
QY 311 AAAGTGGAGAGACAACCTATTTCTTGCAGGGATTGTGCAATTCATCCACATGTGTACTCT 370
Db 311 AAGTGGGGGAACCTACATACTCTCTGCAGAGACTGTGCAGTTGACCCCACTGTGTTTAT 370
QY 371 GTATGGACTGCTTCCAGACAGTGTTCATAAAATCATCTGTTACAAGATGCATCTCTA 430
Db 371 GCATGGAGTGTCTCTCTGGGAAGTATCCATAGAGACCATCGATATAGGATGACCATCGG 430
QY 431 CTGGAGAGGGTCTGTGACTGTGGAGACAGAGGATGGAAGAACTGCCTCTTTTGTG 490
Db 431 GAGGAGGGGCTTCTGTGACTGTGTGACTGTGGAGCGGTGGAAGAGGAGGACCTTACTGCC 490
QY 491 TAAATCATGAACCTGGGAAG---AGCAGGTACTATAAAAAGAGAATTCACGCTGTCCGTTGA 547
Db 491 AGAAGCAACAGCTCAGCAGCTCTCAAGTTTGGAGGAGAGAGATCTCTTTGTGCATCTAT 550
QY 548 ATGAAGAGGTAATTTGTCCAGCCAGGAGAAATATTTTCCTTCAGTGATATAAATATGTCGTAG 607
Db 551 CAGAAGATGTCATCGCCAGAACTTACAACATTTTTCGCTATTATGCTTTTCGATATGCAGTAG 610
QY 608 AAATGACTATATGGGAAGAGGAAAAAGAACTGCGCTCTCTGAACTCCAGATTAAGGKNRY --- 664
```


QY	2753	CTGAGAAGAAGT---AACATTGACTTTTATCATAGGCTTCAAGATTGGGAAGTTGAG	2809	QY	3809	ATATTTCARATAGCATCAAGGAATGGTTATTCTCTTTGCCACACAATTTATAGAATTG	3868
Db	2828	TGGAAGGCACGTGCGAGACCTTACCTTCACACAGAAGATTTCAAAGCCTGCTGATGCAC	2887	Db	3887	CATATTCTGATAGCATAAAAAGAAATGTTAAAGCAATTTGGAAGGCTGCTTACAGGTGG	3946
QY	2810	-----CCATGAATATACAAATGCTTTTGGAAAACTCAAAGGAATTCCTCCAGTTAG	2860	QY	3869	GATTGAAAGTCCACCTGATGAAAGGATCCTCGAGTCCCATCTGCTCACTCGGAGCACCT	3928
Db	2888	CACATACTCCCGAGCATCTAGCTATGCTGGAGACCTTGCAGAACGCCCTCCCTGG	2947	Db	3947	GACTGAAGGTTATCTCTAATGAAGTCAACCCGCTGTGCCATCTCTGCTGGGGACCT	4006
QY	2861	AAGCCAGAGACATGATAACGTGGATCTTACAGATTTTGACACAGTGAAGCATTA	2920	QY	3929	GGCTTTTCACTATCCAGGCAATTTGAAATATTTTGGGAGATGAAGAAACCTCTGTTTG	3988
Db	2948	AAGCCACAGACATGATCAGGTGGTGTCTAAAGATGTTTAATCAATTAAGAAGATA	3007	Db	4007	GTGCATACACCATCCAGAGCATAGAAAGAAATTTTGAGTGATGAGGAAGCCTGTTT	4066
QY	2921	GAGAAAATCTTTTAAATTTAGCAACCATCAGGATCGGAATCTTATTAAGAATGATG	2980	QY	3989	GAGCACTTCAAAATAGCAGCATATGGTCTGAAAGCAATTAATGCACTTTGCACTTGCAC	4048
Db	3008	GAGAGTGTTCATCCAGCAGCCTGTGGCC-----GAGGCGGAGGGAACCAATATGGAGG	3061	Db	4067	GACCTCTGCCTTGTAGACTGGAGACTGTCTCAGTCTGTTAAACAGGTTTGCACAGCAC	4126
QY	2981	AGATTACTCATGATAAAGAAAGCAGAACGAAAAAGAAAGCTCAAGCTGCTAGCTAC	3040	QY	4049	AGAGGATTACCTGTCTCAGTCTGATACAGAAACATCTGGTTCGCTTCTATCAGTTG	4108
Db	3062	AGAGCTCAAGAGCAAGGACAAAGCAGAGAGAAAGAAAGCCGAGATCGCCAGCTGC	3121	Db	4127	ATTGGACAGTGGCGTTACTTCTCTGTGTACAGACACATTTCTGTAAACTCTTTGCATCCT	4186
QY	3041	ATCGCCAGAGCATGATGCTCAGATGCTGCTTACAGAAAAAATTCATTGAAACTCATA	3100	QY	4109	TTCTTCTTCAACATAAATCAGAAGATACACCATGCTTCTCTCTATAGATCTGTTTCATG	4168
Db	3122	CGCGGAGAGATCATGSCCCAGATGCTGAGATCGACGGCACTTCTATTGACGAAACA	3181	Db	4187	TGCTGCTAGTGACAGCTATGAAGACCTCCCGTGCATACATAGACATCGACATGTTCACT	4246
QY	3101	AACCTATGATGACAATACATCAGAAATGCTCGGGAAGAAGATTCATTATGGAAGAG	3160	QY	4169	TTTTGGTGGTGTCTGTGTAGCATTTCCATCTTGTATTTGGGATGACCTGTTGATCTGC	4228
Db	3182	AAGAGCTCTTCCAGCAGACCTAGAGCTGGACACCTCTGCTCTGCCACTCT-----TG	3235	Db	4247	TGCTGGTGGCCTGTGCTGCTTCCAGCTCTGCAGT-----TCAGG	4291
QY	3161	AGAGCAGCCAGCAGTCACTACTCTAGAAATGCTTTGGGTCTTAAACGGGTCCAT	3220	QY	4229	AGCCTTCTCAGTTAGTTCTTCTATTAACACACCTTTATCTCTTCCATTTGATCACCATG	4288
Db	3236	ACAGCAGCCCTCCGCTTTTCAGACGACGCTCTTACAGCAGCTGGGCCAGCAGACACAG	3295	Db	4292	ATTTTTCAGGAAGCAGCTGGCCACTGGGACCTGCGACATCTTCCACTTGGTTACCATGG	4351
QY	3221	CTGTTACTGAAAGAGGTGCTGAGTGCATCTTTGCCAAGAAACAGAGAGGTGA	3280	QY	4289	CACACATGCTTACAGATACTACTTACAGTAGACACAGCCCTACC---CCTTCTCAGGTTTC	4345
Db	3296	TCCCTGAACCCAGACAGTTGTACCTGTATATTATGCAAGAGGAGCAAGAGGTGACTG	3355	Db	4352	CACACATGCTACAGATCTTACTTACCTCATGTACAGAGAGAAATGGCATGGATCAAGAGA	4411
QY	3281	TAGAAAATAATGCCATGATTTATCGGCTGTGTCCAGAAATCTACTGCTTAAACCCAGC	3340	QY	4346	AGAAGACAGTGAAGAGGCTCATTCGCGCATCTTCTTCTTTTGCAGAAATTTCTCAATATA	4405
Db	3356	TGGGACAGCGCGATGCTTGGCAGGCTTGTTCAGAGGTCAACGGTCTGTCAAAAG	3415	Db	4412	ATCCCACTGGGGAAGAAGAACTGGCCATCTCTCTTTTGCACAAAACATTTCAACAGTATA	4471
QY	3341	ACAGGGAAAAACCATAGAACTCTCAGAGAAGCCCTAGACCCATTTTTCATGGATCCAG	3400	QY	4406	CAAGTGGCTCCATTTGGGTGTGATATTCCTGGTGGTATTTGTGGGTCTCACTGAGAAATG	4465
Db	3416	ACAGGACAAAAACCATCGGACCCAGAAAA---TATGATCCATTATTCATGCACCCCG	3472	Db	4472	CTGGAAGTGCCTTGAAGAAGACCCCTCCGCGCTGGCACCTGTGGAGGAGCGCTCGGCGC	4531
QY	3401	ACTTGGCATATGGAACATTATACAGGAAGCTGTGTGATCATGTAATGCAACGAGTGTCTGGC	3460	QY	4466	GCATCACCCCTATCTCGCTGTGCTGCTGTTTCTTCTTCCACTATTTCTTGGGTGAATCTC	4525
Db	3473	ATCTGTCTTGGGACACACACTGSCAGCTGTGGCAGCTTATGATGCCATTTGTTGGC	3532	Db	4532	CCATCATCTTCCCTTCCCAAGTGTCTGCTTTTGTGTTTTCACACTATTTAAATGGAGTCCCG	4591
QY	3461	AGAAGTATTTGAAGCTGTAC-----AGCTGAGCTCTCAGCAGCGCATTCATGTTG	3511	QY	4526	CGCTCAGGAACCTGCATACCAATTTCTGCAAGAGAGAGTACAGTGCACCTCTGTAGCTATC	4585
Db	3533	AAAGGTATTTGATTCCTGTTCAAGCAAGGAGCAGCAGGAGGAGCGGCTGCGCTTGC	3592	Db	4592	CCCTCCAGACCTTCA---AGTTTCTGGAACAAGCCATTTTGAACACTTATGTAACATACC	4648
QY	3512	ACCTT---TTTGACTTGGAAAGTGGAGAAATCTTTGGCCCTCTTTGCAAACTCTGTGTGA	3568	QY	4586	TATCTTTACTACAAATTTGTTCTCTCTTCCAGGAATATTTGGGATACTGTAGAGGCCCT	4645
Db	3593	ACACTAGCTACGATGATAGAAATGCGAGTTCTCTGCCCCGCTCTGTGAGTGCCTGAGCA	3652	Db	4649	TTTCCCTACCAACCAACCTCATTCACCTTTTCAAGAAAAACAGTGACATCATGAATCCC	4708
QY	3569	ATACTGTATCCCATTAATCTTTGCAACCTCAAAAGATAAACAGTGAAGATGAGATG	3628	QY	4646	TGCTCAGAGCGGTGTGCAGATCCTGCTTACTAACTGTTTGAAGCAAAAAACACCG	4705
Db	3653	ACAGGGTATCCCC-----CTGCTGCTTCTCCAGGAGCATCTCAGCAGGAGGTTAA	3706	Db	4709	TGATTGAAATTTGGTGCCAGAACAGTGAAGTTAAACGGTATCTTAATTTGGGAGAGAGAG	4768
QY	3629	CTCTTGCTCACTTTTGACCTGGCAGGTGGATACAGACTCTTCTTGGCCAGAAATACAG	3688	QY	4706	TGCTCAGGTACCTCAGAAAAAGAAATAGTTTGTATAGAGCTTCTCTGATGACTATAGCTGCC	4765
Db	3707	ATTTTTCAGACCAACCATCTGGCACAGTGGACAGAGCAGTAACACAGCAGATAAAGG	3766	Db	4769	CGATAAGCTTACCCACAGGAGCTAACAACTGATAGACCTTCCAGAGGATTTACAGAGCC	4828
QY	3689	GTTATAATTAAGACATCTCAAGGAGAAAAACCAATCTTATTTCTTAATCAAGGAA	3748	QY	4766	TCTGAATCAAGTCTCTCATTTTCAGGTGCCACGGCTGCAGATGATGAGCGAAAGCATC	4825
Db	3767	TGGTCCAGATGCTGAGGAGAAAGCAATGTGCTGTACACGCTCTTCTTCAGAGGACACAG	3826	Db	4829	TCATTAAACAGCATCCAATTTCTCGTGCCCCAAATCAGGTGGGCGACAAGAGCAGAGCTC	4888
QY	3749	TGGGAGATTTACTTTTGAGTTGCCATTCATCTGAGTTTGGCTGTGAGTCTTCGATTA	3808	QY	4826	CTGCTCTGCTCTTCTGTGGGCTTACTATGTTCTCAGAACATTTGCTGCCAGGAAA	4885
Db	3827	AAGCCATGAATATAATACCGATCCCGAAGGCTTCAGGCTGATTTTATCTCAGGAACC	3886	Db	4889	CTACTCTGCTCGTGTGTGGAGTCTCCTCTGCTCTCAGATTTACTGTCTGCCAAGCTG	4948
				QY	4886	TTGTGAACGGGGAAGAGTTGGAGCTTGCACTTTTTCAGGCATTTTCTCAGTGTGGACCGGAG	4945

```
Db 4949 AGCTGGAGGTGAGGACGTCGGAGCTGCACAGCACACACTTACTCCTCGGCTCGGGG 5008
QY 4946 TCTGCATTTTCCATAAAATCAGAGATGCCGAGTGGTCTGTTGAAGGTAAAGCCAGAG 5005
Db 5009 CCGGCATCTTCTGAGAGTGGCGGAATGTCAAGTGTATTTTGTAGTGGCAAAACAAAG 5068
QY 5006 GCTGTGCCATATCCAGCTCTTACTTGGATGAATATGGAGAAACAGACCTGGCTGAAGA 5065
Db 5069 GATGTTTTTATTTCTCTCTTACCTTGACGACTATGGAGAGACCGACAGGGACTCAGAC 5128
QY 5066 GGGCAACCCCTTCAATTTATCTGTGAGCGGTATCGGAAGCTCCCAATTTGGTCTGCGAAC 5125
Db 5129 GAGGAATCCTTTACATTTATGCGAAGACGGTTTCGAAGATCCAGAGCTCTGGCAGC 5188
QY 5126 AACACTGCAATATAGAAGAGATTTGCTAGGAGCCAGAGACTAATCAGATGTTATTTGGAT 5185
Db 5189 AGCATAGTATCAGAGGAGATCGGACACGCGCAGGAGGTAAACAGACCTTGGTCGGAA 5248
QY 5186 TCAACTGGCAG 5196
Db 5249 TTGACTGGCAG 5259

RESULT 12
AR030785
LOCUS AR030785 1001 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 2 from patent US 5861312.
ACCESSION AR030785
VERSION AR030785.1 GI:5943999
KEYWORDS
SOURCE
ORGANISM
REFERENCE
  1 (bases 1 to 1001)
AUTHORS Varshavsky,A. and Kwon,Y.Tae.
TITLE Nucleic acid encoding mammalian UBrl
JOURNAL Patent: US 5861312-A 2 19-JAN-1999;
FEATURES
  Location/Qualifiers
    source
      1..1001
        /organism="unknown"
BASE COUNT 363 a 186 c 205 g 247 t
ORIGIN
```

```
Query Match 19.1%; Score 996.2; DB 6; Length 1001;
Best Local Similarity 99.7%; Pred. No. 4.9e-232;
Matches 998; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2059 ATGGATCCCAATAAGTCTTGTGTACTTGGTACTTCAGAGGTATGAACCTGCGGAGGCTTTT 2118
Db 1 ATGGATCCCAATAAGTCTTGTGTACTTGGTACTTCAGAGGTATGAACCTGCGGAGGCTTTT 60
QY 2119 AACAAACCATATCTACAAAAGACCGAGGATTTGATTAACAATATAATACACTAATAGAA 2178
Db 61 AACAAACCATATCTACAAAAGACCGAGGATTTGATTAACAATATAATACACTAATAGAA 120
QY 2179 GAATGCTTCAGTCTCTCATCTATATTGTGGTGAGCTTATGCTTACCTGAGTGGGAAT 2238
Db 121 GAATGCTTCAGTCTCTCATCTATATTGTGGTGAGCTTATGCTTACCTGAGTGGGAAT 180
QY 2239 GTGACCAAAAGAGAGGTCACAATGAGAGAAATCATTACCTTGTGATTTGAACCCATG 2298
Db 181 GTGACCAAAAGAGAGGTCACAATGAGAGAAATCATTACCTTGTGATTTGAACCCATG 240
QY 2299 CCACAGTGCCTATTCGCAAAATTTTACCTGAGAAATGAATAATGAACCTGGCTTAGAG 2358
Db 241 CCACAGTGCCTATTCGCAAAATTTTACCTGAGAAATGAATAATGAACCTGGCTTAGAG 300
QY 2359 AATGTCTAAACAAAGTGGCCACATTTAAGAAACCGAGGTATCAGGCCATGGAGTTTAT 2418
Db 301 AATGTCTAAACAAAGTGGCCACATTTAAGAAACCGAGGTATCAGGCCATGGAGTTTAT 360
QY 2419 GAACTAAAAGATGAATCACTGAAAGACTTCAATATGTACTTTTATCATTTACTCCAAAACC 2478
```

```
Db 361 GACTAAAAGATGAATCACTGAAAGACTTCAATATGTACTTTTATCATTTACTCCAAAACC 420
QY 2479 CAGCATAGCAAGGCTGAACATATGCAAGAAAAGAGAGAAAACAAGAAAACAAGATGAA 2538
Db 421 CAGCATAGCAAGGCTGAACATATGCAAGAAAAGAGAGAAAACAAGAAAACAAGATGAA 480
QY 2539 GCATTTGCCGCCACACACACCTCTCGAATTTCTGCCCTGCTTTTCAGCAAAAGTGAATTAACCTT 2598
Db 481 GCATTTGCCGCCACACACACCTCTCGAATTTCTGCCCTGCTTTTCAGCAAAAGTGAATTAACCTT 540
QY 2599 CTCAACTGTGATATCATGATGATCATCTCAGGACCGGTATTTGAGCGGGCAATTAACACA 2658
Db 541 CTCAACTGTGATATCATGATGATCATCTCAGGACCGGTATTTGAGCGGGCAATTAACACA 600
QY 2659 GATTCTAACTTTGGGACCGAAGGATGCTCCAAATGGCTTTTCATATTTCTGGCATTTGGGT 2718
Db 601 GATTCTAACTTTGGGACCGAAGGATGCTCCAAATGGCTTTTCATATTTCTGGCATTTGGGT 660
QY 2719 TTACTAGAAGAGAACACAGACTTCAAAAGCTCCTGAAGAAAGTAAACATTTTGACATTT 2778
Db 661 TTACTAGAAGAGAACACAGACTTCAAAAGCTCCTGAAGAAAGTAAACATTTTGACATTT 720
QY 2779 TATCATAAAGCTTCAAGATTGGGAAGTTTCAAGCATGAATATACAAATGCTTTTGGAAAAA 2838
Db 721 TATCATAAAGCTTCAAGATTGGGAAGTTTCAAGCATGAATATACAAATGCTTTTGGAAAAA 780
QY 2839 CTCAAAAGGAATTTCCCAAGTTAGAAGGCCAAGGACATGATAAGCTGGATCTTCAGATG 2898
Db 781 CTCAAAAGGAATTTCCCAAGTTAGAAGGCCAAGGACATGATAAGCTGGATCTTCAGATG 840
QY 2899 TTTGACACAGTGAAGCGATTAAAGAGAAAAATCTTTGTTAAATTTAGCAACCATCAGGA 2958
Db 841 TTTGACACAGTGAAGCGATTAAAGAGAAAAATCTTTGTTAAATTTAGCAACCATCAGGA 900
QY 2959 TCGGAATCTATTAGAATGATGAGATTACTCATATAAGAAAAGAGAGAGAGAAAAAGA 3018
Db 901 TCGGAATCTATTAGAATGATGAGATTACTCATATAAGAAAAGAGAGAGAGAAAAAGA 960
QY 3019 AAGCTGAAGCTGTAGCTACATCGCCAGAAAGATCATGGC 3059
Db 961 AAGCTGAAGCTGTAGCTTCATCGCCAGAAAGATCATGGC 1001

RESULT 13
AR121464
LOCUS AR121464 1001 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 2 from patent US 6159732.
ACCESSION AR121464
VERSION AR121464.1 GI:14105040
KEYWORDS
SOURCE
ORGANISM
REFERENCE
  1 (bases 1 to 1001)
AUTHORS Varshavsky,A. and Kwon,Y.Tae.
TITLE Nucleic acid encoding mammalian Ubr1
JOURNAL Patent: US 6159732-A 2 12-DEC-2000;
FEATURES
  Location/Qualifiers
    source
      1..1001
        /organism="unknown"
BASE COUNT 363 a 186 c 205 g 247 t
ORIGIN

Query Match 19.1%; Score 996.2; DB 6; Length 1001;
Best Local Similarity 99.7%; Pred. No. 4.9e-232;
Matches 998; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2059 ATGGATCCCAATAAGTCTTGTGTACTTGGTACTTCAGAGGTATGAACCTTGGCAGGCTTTT 2118
Db 1 ATGGATCCCAATAAGTCTTGTGTACTTGGTACTTCAGAGGTATGAACCTTGGCAGGCTTTT 60
QY 2119 AACAAACCATATCTACAAAAGACCGAGGATTTGATTAACAATATAATACACTAATAGAA 2178
```

[illegible]

Db 361 GAACATAAAGATGAATCACTGAAAGACTTCAATATATGTACTTTTATCATTTACTTCCAAAACC 420

Qy 2479 CAGCATAGCAAGCGTGCACATATGCAGAGAAAGAGAAACAGAAACAAAGATGAA 2538

Db 421 CAGCATAGCAAGCGTGAACATATGCAGAGAAAGAGAAACAGAAACAAAGATGAA 480

Qy 2539 GCATTGCCGCCACCAACCACTCCTCTGAATTCCTCCCTGCTTTTCAGCAAAAGTGATTAACCTT 2598

Db 481 GCATTGCCGCCACCAACCACTCCTCTGAATTCCTCCCTGCTTTTCAGCAAAAGTGATTAACCTT 540

Qy 2599 CTCACATGTGTATCATGTATGATATCTCTCAGGACCGTATTTGAGCGGCAATAGACACA 2658

Db 541 CTCACATGTGTATCATGTATGATATCTCTCAGGACCGTATTTGAGCGGCAATAGACACA 600

Qy 2659 GATTCTAAGTGTGGACCGAAGGATGCTCCAAATGGCTTTTCATATTCGCGATTTGGGT 2718

Db 601 GATTCTAAGTGTGGACCGAAGGATGCTCCAAATGGCTTTTCATATTCGCGATTTGGGT 660

Qy 2719 TTACTAGAGACAAGCAACAGCTTCAAAAAGCTCTCGAAGACGTAAGTAACTTTGACTTT 2778

Db 661 TTACTAGAGACAAGCAACAGCTTCAAAAAGCTCTCGAAGACGTAAGTAACTTTGACTTT 720

Qy 2779 TATCATAAGCGTTCAAGATTTGGAAGTTCAGGCATGAATATACAAATGCTTTTGGAAAAA 2838

Db 721 TATCATAAGCGTTCAAGATTTGGAAGTTCAGGCATGAATATACAAATGCTTTTGGAAAAA 780

Qy 2839 CTCAAAGGAATTTCCCGAGTTAGAACGCCAGAGGACATGATTAACGTTGGATCTTCAGATG 2898

Db 781 CTCAAAGGAATTTCCCGAGTTAGAACGCCAGAGGACATGATTAACGTTGGATCTTCAGATG 840

Qy 2899 TTTGACACAGTCAAGCGATTAGAGAAAAATCTTGTTTTAAATGTACCAACCATCAGGA 2958

Db 841 TTTGACACAGTCAAGCGATTAGAGAAAAATCTTGTTTTAAATGTACCAACCATCAGGA 900

Qy 2959 TCGGAATCTATTAAAGAAATCATGAGATTACTCATGATAAAGAAAAACAGAACCAAAAAAGA 3018

Db 901 TCGGAATCTATTAAAGAAATCATGAGATTACTCATGATAAAGAAAAACAGAACCAAAAAAGA 960

Qy 3019 AAAGCTGAAGCTGCTAGGCTATACATCGCCAGAGATCATG 3057

Db 961 AAAGCTGAAGCTGCTAGGCTTTCATCCGAGAGATCATG 999

RESULT 15

AB002347

LOCUS

AB002347 Human mRNA for KIAA0349 gene, partial cds. 6158 bp mRNA linear PRI 06-OCT-2001

DEFINITION

AB002347

VERSION

AB002347.1 GI:2224638

KEYWORDS

KIAA0349.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

REFERENCE

1 Nagase,T., Ishikawa,K., Nakajima,D., Ohira,M., Seki,N., Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O. Prediction of the coding sequences of unidentified human genes. VII. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro

AUTHORS

Ohara,O., Nagase,T., Kikuno,R. and Nomura,N.

TITLE

Direct Submission

JOURNAL

Submitted (28-MAR-1997) Osamu Ohara, Kazusa DNA Research Institute; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan

MEDLINE

97349984

PUBMED

9205841

REFERENCE

2 (bases 1 to 6158)

AUTHORS

Ohara,O., Nagase,T., Kikuno,R. and Nomura,N.

TITLE

Direct Submission

JOURNAL

Submitted (28-MAR-1997) Osamu Ohara, Kazusa DNA Research Institute; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan

FEATURES

source

1. .6158

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="HG1561"

/sex="male"

/tissue_type="brain"

/clone_lib="pBluescriptII SK plus"

1. .6158

/gene="KIAA0349"

<1. .3828

/gene="KIAA0349"

/codon_start=1

/protein_id="BAA20806.1"

/db_xref="GI:2224639"

/translation="LKVLISKPTSEWDELROKFLGDFLEKLLKCMQGMDDPTTROV

GOHIEPEHEAFTLQMKLTHVISMQDMCASDEKVLIEAYKCLAVLMQCHGYTD

GEOPITLISCGHSVETIRYCVSQKSIHLPSRLLAGLHVLLSKSEVAYRFPPELLP

SELSPMLEIHLRLMLSLRPETYQIFSTPDYGRFSEITHKDVQONNLIEMLYL

QTVGSMDPNHFLMILSRPELYQIFSTPDYGRFSEITHKDVQONNLIEMLYL

IIMLGERFSPGVQVNDIEIKREIITQLSIKPMASELVSLEPDKENKGTGSESVI

EVAHFKKPGLTGRGWYELKPECAKENLYFYHFSRAEQSKAEAROKLKRQNRDPA

LPPVLPFCPLFASLUNILQSDVLMCIIMGITLQWAVEHNGYANSESMQRLVHLIGM

ALQEEKOHLENVTEHVVFTFTQKISKPEAGPKNSPILAMLETLQNAIPYLEVHKDM

IRWLTFNVAKKMRESPTSPAETETIMESRDKRAERKKAIEARLRREKIM

AOEMORHFIIDENKELFOOTLELDASTAVLDHSPVASMNTLALGPTQVPEQORQ

FVTCILCOEEQEVKVESRAVLAAFVORSTVLSKNRSKFIODPKYDPLFMPHPLDSCG

THTSSCGHIMHACWORYFDSVQAKEQRROORLRLHTSYDVENGEFFLCPCECLNTV

IPLLPRNIFNNRNFSDQPNLWRTISQKIKALQFLRKEESTPNNASTKNSENV

DELQLPGRFPDRPKIPYSESIKEMLTTFGTATYKVLKVHPNEEDRPVIMCWGSC

AYTQSTIERISDEDKPLFGPLPCRLDDCLRSRFAAAHWTVASVSVOHFKCLFA

SLVPSNDSEELPCITLIDMFHLVLGLVLAFLALQODPSGISLGTDLHIFLVTMAH

IIOILLTSCTEENGMDENPPCEESAVLALYKTHOYTGSALEKEIPSGHWLRSVRA

GIMPFKCSALFFHYLNGVSPDPDQVPGTSHFEHLCSYLSLPNNLILCLFOENSEIMN

SLIESWCNRSVRYLEGERDAIRYPRESNKLINLPEDYSLINQASNFSCPCKGGDK

SRAPTLCLVCGSLSCYCCYCCOEGEDGACTAHTYSCGSGVGIPLURRECVQLFL

AGTKGCFYSPYLDYDGETDQGLRRGNPLHLCKERFKIKQLKWHQHSVTEIGHAQE

ANOTLVGIDMQHL"

BASE COUNT 1832 a 1199 c 1320 g 1807 t

ORIGIN

Query Match 18.8%; Score 979.4; DB 9; Length 6158;

Best Local Similarity 55.9%; Pred. No. 6.1e-228;

Matches 2156; Conservative 0; Mismatches 1611; Indels 90; Gaps 12;

Qy 1390 CTAAAGTATATCTCGATCAGCAAAACCACCAATATGACAGAAAGATTAAGATGCAGTTC 1449

Db 1 CTCAAGTATGTGTTAATTAGCAAAACCACTGAATGGTCAGATGAGCTCAGGCAGAAAGTTC 60

Qy 1450 CTTCAAGGTTTTCGATCTTTTTTGAAGATTTCTTACCTGTATGTCAGGGAATGAAGAAATC 1509

Db 61 CTAGAAGGTTTGTATGCTTTTTTGAATTTACTAAATGTATGTCAGGGAATGGATCCAATT 120

Qy 1510 CGAAGACAGGTTGGGCAACACATGGAAGTGGATCCTGATTTGGGAGGCTGCCATTGCTATA 1569

Db 121 ACAGCTCAAGTAGGACACATATTGAATGGAACAGAGATGGGAAGCAGCCCTTCACACTA 180

Qy 1570 CAGATGCAATGAAGAATATTTTACTCATGTTTCCAAAGAGTGGTGTGCTTGTGATGAGAA 1629

Db 181 CAAATGAAATTAACACATGTCATTTCAATGATGTCAGGACTGTGTGCTTCAGATGAAAA 240

Qy 1630 CTCCTTACTTGGGCTTATAAAGAATGTCACAAAGCTGTGATCAGGTCGAGTACCAGTTTC 1689

Db 241 GTGTTAATCGAAGCTTACAGAAATGTCTCGCTGTACTGTCAGTGTGATGGTGGTTAT 300

Qy 1690 ATATCTAGTAGCAAGACAGT---AGTACAATCGTGTGGACATAGTTTGGAAAAACAAGTCC 1746

Db 301 ACTGATGGTGAACAGCCCAATCACACTAAGCATTTGTGGACATTCAGTGGAAACTATCAGA 360

Qy 1747 TACAGAGTATCTGAGGATCTTTGTGAAGCATATCATCTGCCACTCTCTAGGACCCCTTGTGTT 1806

Db 361 TACTGTGTTTCCCAAGAAAAAGTTAGCATTCACCTCCCAAGTTTCTCGCTTACTTGCAGGT 420

Qy 1807 CTTTCATGTACGTTTAAAGCAGCGTGGGTGCTGCTTTTCAAGACTGCATGAATTTTGTGCTTTT 1866

Db 421 TTACATGTATTATTAAAGCAAAAGTGGCATATAAATTTCCAGAGCTCCTCACTCTA 480

OY 1867 GAGGACTTCAAGTAGAGTACTAGTGAATCCTTTACGTTGCTGCTGTTGCC 1926
 DB 481 AGTGAACCTTAGCCGCCGCCATGTTGATAGAACCCCTTAGAGTGTCTGCTGCC 540
 OY 1927 CAGGTGTTGCTGAGATGTGGCGAAGAAATGAGACTGCTTATATGAGCCAGGCTTTAT 1986
 DB 541 CAAGTACATGCCGGAATGTGAGAGAAATGGTCTCTCTAGTAACCAAGATTTATTAC 600
 OY 1987 TACCAAGATGTTAATGTCAGAGAAAGAAATGTATGATAAGATATCATGCTTCAGATT 2046
 DB 601 TACCAATATGTAAATGACAGAGTATGATGTTGACAAAGATGATATGATGCTTCAGACA 660
 OY 2047 GGTGATCTTTAATGATGCCATATAGTCTGTTACTGTTACTGATCAGAGTATGACT 2106
 DB 661 GGTGCTCATATGATGATCCAAATCATTTCTGATGATCATGCTCAGCCGCTTTGACTT 720
 OY 2107 GCCGAGGCTTTTACAGACCATATCTACAAAA-----GACGAC 2145
 DB 721 TATCAGATTTTCAAGTCCAGACTATGAAAAAGATTAGTTCTGAGATTACCCATTAG 780
 OY 2146 GATTGATTTAAACAATATATATCACTAATAGAGAAATGCTTCAAGTCTCATATAT 2205
 DB 781 GATGTTGTCAGACAGAACATCTCTAATAGAAATGCTATACCTCATATATATGCTT 840
 OY 2206 GTGGGTGAGCGTTATGTAACCTGAGTGGAAATGTGACCAAGAGAGTCAATAGACA 2265
 DB 841 GTGGAGAGAGATTTAGTCTGAGAGTGGACAGGTAATGTACAGATGAAATCAAGCCA 900
 OY 2266 GAAATCATTCACCTTCTTTCATTAACCCATGCCACAGATGCCATTCACCAAAATTTA 2325
 DB 901 GAGATTTATCCATCACTTATGATATCAAGCCTATGCTCATATGATGATTTGTTAAATCTT 960
 OY 2326 CCTGAGAAATGAAATATATGAACTGCTTGAAGATGTCATTAACAAAGTGGCCACATTT 2385
 DB 961 CCTGAGAGATGAGAACAGAGAGACTGGAGAGTGTATGCAAGCAGTGGCCATTTTC 1020
 OY 2386 AAGAAACAGAGTGTACAGAGCCATGAGTTTATGAATTTAAAGATGATCACTGAAAGC 2445
 DB 1021 AAGAAACCTGATTTAAACAGAGAGGAGCATGTATGAACTGAAACAGAAATGCCAAAGAG 1080
 OY 2446 TTCAATATGACTTTTATCATTTACTTCCAAAACCCAGCATAGCAAGGCTGACATATGACG 2505
 DB 1081 TTCAACTGTATTTTATCATCTTTTCAAGGCGAGACAGTCCAAAGCGACAGAAAGCGCAA 1140
 OY 2506 AAGAAAGAGAGAAACAAAGAAACAAAGATGAGCATTTGCCGACACACCTCTCTGAA 2565
 DB 1141 CGAAATTTGAAAGACAAATATAGAAATACAGCACTCCACCTCGGTGTTGCCCTCA 1200
 OY 2566 TTCTGCCCTGCTTTCAGCAAGATGATTAACCTTCAACTGTATATCATGATGATTT 2625
 DB 1201 TTCTGCCCTGCTTTCAGCAAGCTGTTAACTTTGACGTCAAGTGTATGTTGTCATC 1260
 OY 2626 CTCAGAGCCGTTATTTGAGCGGCATATAGACACAGTTTCACTTTGAGCCGCAAGGATG 2685
 DB 1261 ATGGGAACAATTTGCAATGGCGTGTGAACAATATGATGATGCTGTCGCAAGTCCATG 1320
 OY 2686 CTCCAATATGCTTTTCAATTTCTGSCATTTGGTTTACTAGAGAGAGCAACAGCTTCA 2745
 DB 1321 CTGCAAAAGGTTGTTATTTAATTTGCGATGGCATACAGAAAGAAACAAATTTAGAG 1380
 OY 2746 AAAGCTCTGAG---AAGAAATACATTTGACTTTTATCATAGGCTTCAAGATTTGGA 2802
 DB 1381 AATGTCAGAGAGAGCATGTAGTACATTTTACCTTCACTCAAGAAATATCAAAACCTGCT 1440
 OY 2803 AGTTGAGCATGAAT-----ATACAATGCTTTTGGAAAACTCAAGAAATTCCTC 2853
 DB 1441 GAAGCGCAAAAAATTTCTCTAGCATGATGATGATGTAAGCAACATAAATGCTCTCC 1500
 OY 2854 CAGTTAGAGGCGAGAGAGATGATTAAGCTGATGATGATGATGATGATGATGATGATG 2913
 DB 1501 TACCTAGAAGTCCAAAGACATATGCTGGGATGATGATGATGATGATGATGATGATG 1560

OY 2914 CGATTAAGAAAAATCTGTTTAATTTGATACCAACCATCAGATGCGAATCTATTAAG 2973
 DB 1561 AAGATGAGGAGA-----GTTCACCTACAGTCCCGGCGAGAGACAGAAAGACATA 1614
 OY 2974 AATGATGAGATTTACTGATTAAGAAAAACAGACCAAGAAAGAAAGAAAGTGAACCTCT 3033
 DB 1615 ATGGAAGAGATTTCAAGGCAAAAGACAAACCTAGAGGAAAGAAAGAAAGAGATTTGCC 1674
 OY 3034 AGGCTATCGCCAGGAAGATCAGTGGTATGCTGCTTACAGAAAACTTCATGAA 3093
 DB 1675 AGACTGGCAGAGAAAAATCATGCTCAGATGCTGAAATGACAGCGCATTTATGAT 1734
 OY 3094 ACTCATTAATCATGATGATCAATACATCAAGAAATGCTGGGAAAGAAAGATTCATG 3153
 DB 1735 GAAACAAAGAACTTTTACAGACATTTAAGCTGATGCTCAACCTGCTGCTT 1794
 OY 3154 GAGAGAGAGACACCCAGCAGTGTGACACTCTGAAATGCTTGGTCTTAAACGG 3213
 DB 1795 GATCA-----TAGCCCTGTGGCTTACATATGACACTTACAGCACTGGTCCACACA 1848
 OY 3214 GGTCCATCTGTACTGAAAAAGAGTGTGACATGCTTGGCCAAAGAACAGAG 3273
 DB 1849 ACTCAGGTTCTGACAAAGACATTTGTTACATGATATATGTTGTCAGAGAGACAGAA 1908
 OY 3274 GTGAAATTAAGAAATATGCTCATGTATATGCGCTGTGCGAAGAAATCTACTGCTTA 3333
 DB 1909 GTTAAAGTGAAGACAGGCAATGCTTGGCACATTTGTTACAGATCACTGATATTA 1968
 OY 3334 ACCAGACAGAGGGAACCAATGAACTGACAGAGAGAGCCATAGACCCTTTTCAG 3393
 DB 1969 TCAAAAAACGAAATTAATTTTCAAGATCCAAAAAAAT---GATCCATATTTCA 2025
 OY 3394 GATCCAGACTTGCATATGAACTTATACAGAGAGTGTGCTCATGTAATGACAGCATG 3453
 DB 2026 CACCCGATGCTGCTTGTGGAACACACATGATGCTGTGGGACATATGATGATGCCAT 2085
 OY 3454 TGTGCGCAGAAATTTTGAAGCTGTGACAGTGTGCTGACAGCGCATCATGTTTAC 3513
 DB 2086 TGTGTCMAAGATTTTGAATTTGATTCGTTCAAGCTTAAGAACAGCAAGCAACAGATTA 2145
 OY 3514 CTTT-----TGACTTGGAAAGTGGAGAAATATCTTGGCCCTTTGCAAACT 3561
 DB 2146 CGCTTACATACGAGCTATGATGTAAGAAACGAGAAATCTTTGCCCCCTTTGGAATGC 2205
 OY 3562 CTGTGCAATACCTGTGATCCCATTTATCTTTGCAACCTCAAAAGATTAACAGTGAAT 3621
 DB 2206 TTGAGTAATATCTGTATATTC-----TCTGCTGCTCTCCAAAGAAATTTTAAACAC 2259
 OY 3622 GCAGATGCTTGTCTCAACTTTTACCTGGACGCTGATGATACAGACTGTTTGGCCAGA 3681
 DB 2260 AGGTTAAATTTTTCAGACCAACCAAAATCTACTCAGTGAATTAACAATATGTCAGCA 2319
 OY 3682 ATATACGTTATATATTAACACATGTAAGAGAAACCAATTCATTTCTTTAT 3741
 DB 2320 ATTAAGACATTAAGTTTCTTAGAGAAAGAAAGTCTCTTAATTAATGCTCTACAAAG 2379
 OY 3742 CAAGGAATGGAGATTTACTTTGAGTTCATTCATTCCTGATTTTGGCTGAGACT 3801
 DB 2380 AATTCAGAAATATGATGATTAATTAAGCTCTCCCAAGGTTCCAGGCTGATTTTCTGCT 2439
 OY 3802 TCGATTAATATTTCAATATACATCAAGAAATGTTATCTTTGCCAACAATATAT 3861
 DB 2440 AAGATCCCTATTTCTGAGACATTAAGAAATGCTAAGCATTTTGAATGCTACCTAG 2499
 OY 3862 AGAATTTGAAATGTCACCTGATGAAAGAGATCTGAGATGCCATCTGACCTG 3921
 DB 2500 AAGTGGAGCTAAAGGTTTCAATCCCAATGAAGAGATCTGTGTTCCCATATGTTGG 2559
 OY 3922 AGCAGCTGGCTTCTACTATCCAGGCAATGAAATATATTTGGAGGAAGAAACCT 3981
 DB 2560 GGTAGCTGGCGTACACCATCAAGCATGAAAGAAATTTTGAATGATGAAGATTAACCA 2619
 OY 3982 CTGTTTGAGCACTTCAAAATAGCAGCATATGCTGTGAAGCATTAATGCACTTGCA 4041

Db 2620 TTGTTTGGCCCTTACCTTGCAGACTGGATGACTGCTTAGGTCATTTGACGAGATTTGCC 2679
QY 4042 GTTGACACAGGATTTACCTGCTCAGGTCCTGATACAGAAACATCTGGTTCCTCTCTA 4101
Db 2680 GCAGCACACTGGACAGTCAGTTTCAGTGGTGCAGAGACATTTTGTAAACTTTT 2739
QY 4102 TCAGTTGTTCTTCCCTAACATAAAATCAGAAATACACCATGCTCTCTCTATATAGTCG 4161
Db 2740 GCATCACTGGTCCCTAATGACAGCCATGAGAACTTCCATGTCATATATAGATTTGACATG 2799
QY 4162 TTTTCATGTTTGGTGGGCTGCTGTGTAGCATTCCTCCATCTGTTGATTTGGGATGACCCCTGTT 4221
Db 2800 TTTTCATTTAATGGTGGGCTTGGTGTGCTTGCATTTCCCTGGCTTGCAGTG----- 2846
QY 4222 GATCTGCAGCCTTCTTTCAGTTAGTTCTTCTCTATTAACACCTTTATCTCTTCCATTTGATC 4281
Db 2847 --TCAGGATTTTTCAGGATCAGCCTTGGCACTGGAGACCTTCACATTTTCCATCTGGTT 2904
QY 4282 ACCATGGCACACATGCTTCAGATACCTTACAGTACAGTACAGAGGCC---TACCCCTTGGT 4338
Db 2905 ACTATGGCACACATCAGATCTTACCTACCTCATGTACAGAGAGATGGCATGGAT 2964
QY 4339 CAGGTTCAAGAGACAGTGAAGGCTCATTCGCGATCTTCTTTTCGAGAAATTTCT 4398
Db 2965 CAAGAAATCCCCCTTGTGAAGAAGAAATCAGCAGTTCTTGTCTTGTATAAACACTTCAC 3024
QY 4399 CAATATACAAAGTGGCTCCATTTGGGTGTGATATTCCTGGCTGGTATTTTGGGTCTCACTG 4458
Db 3025 CAGTATACGGGAGTGCCTTGAAGAAATACCATCCGCTGGCATCTGTGGAGGAGTGC 3084
QY 4459 AAGAATGGCATCACCCCTTATCTTCCGTGTGTGCAATTTGTTTCCACTATTTTACTTGGG 4518
Db 3085 AGAGCTGGAATCATGCTTCTCCTGAAAGTGTCTGCTTTATTTTTCATTTAAATGGA 3144
QY 4519 GTAACCTCCGCTGAGAACTGCATACCAATTCGAGAGAGGAGTACAGTGCACCTGT 4578
Db 3145 GTTCCCTTCCCCACCCGACATCTCA---AGTTCTTGGAAACAGCCATTTTGAACATTTATGT 3201
QY 4579 AGCTATCTATCTTACCTACAAATTTGTTCTCTCTCTCCAGGAATATTGGGATCTGTA 4638
Db 3202 AGCTATCTTCCCTACCAACACCTCATTTGCTTTTCAAGAAATAGTGAGATATG 3261
QY 4639 AGCCCTTGTCTCAGAGGGGTGTGAGATCTTCCCTTACTAACTGTTTGAAGCAAAA 4698
Db 3262 AATTCACGTATGAAAGTTGGTCCGCTAACAGTGAAGTTAAAGATATCTAGAAGGTGA 3321
QY 4699 AACACCGTGTGAGTACCTAGAAAAGAAATAGTTTGTATAGAGCTTCTGTAGTACTAT 4758
Db 3322 AGAGATGCTATAAGATATCCAGAGAACTTAACAAATTAATAACCTTCCAGAGGATTAC 3381
QY 4759 AGCTGCTCTCTGAATCAAGCTTCTCATTTTCAGTGGCCAGCGTCTGCAGATGATAGCGA 4818
Db 3382 AGCAGCTCATTAATCAAGATCCATTTCTGCTGCCGAAATCAGGTGGTGTATAGAGC 3441
QY 4819 AAGCATCTGTCTCTGCTCTTCTGTTGGGGCTATATATGTTCTCAGAAACATTTGCTGC 4878
Db 3442 AGAGCCCAACTCTGTGCTTGTGTGGGATCTCTGCTGTGCTCCAGAGTTACTGCTGC 3501
QY 4879 CAGGAAATTTGTAACGGGGAGAGGTTGGAGCTTGCATTTTTCACGCACTTCACTGTGGA 4938
Db 3502 CAGACTGAACTGGAAGGGGAGGATGAGGAGCCTGCACAGCTCACTCTCTGTGGC 3561
QY 4939 GCGGAGTCTGCAATTTTCTTAAATCAGAAATGCGAGTGGTCTCTGTTGAAGGTAAA 4998
Db 3562 TCTGGAGTGGGCATCTTCTCAGAGTACGGGAATGTCAGGTGCTATTTTAGCTGGCAA 3621
QY 4999 GCCAGAGGCTGTGCTCTATCCAGCTCTTACTTGGATGAATATGGAGAAACAGACCCCTGGC 5058
Db 3622 ACCAAGGCTGTTTATTTCTCTCTTACTTGTGATGACTATGGGAGACCGACAGGA 3681
QY 5059 CTGAAGAGGGGCAACCCCTTCAATTTATCTGCTGAGGGGTATCGGAAGCTCCATTTGGTC 5118

Db 3682 CTCAGACGGGAAATCCTTTTACATTTTATGCAAAGAGCGATTCAAGAAGATTTCAGAAGCTC 3741
QY 5119 TGGCAACAACACTGCTCATTTATAGAAGAGATTCTCTAGGAGCCCAAGAGACTAATCAGATGTTA 5178
Db 3742 TGGCACCAACACACTGTTCACAGAGGAAATTTGGACATGCACAGGAGCCCAATCAGACACTG 3801
QY 5179 TTTGGATTCAACTGGCA 5195
Db 3802 GTTGGCATTGACTGGCA 3818

Search completed: September 27, 2003, 09:29:53
Job time : 12271.1 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2003, 19:32:57 ; Search time 820.105 Seconds

(without alignments)
17132.652 Million cell updates/sec

Title: US-09-724-126A-18

Sequence: 1 atggcgagagagagagctgtg.....tcaactgagcagctactgtga 5205

Scoring table: IDENTITY_NDC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 segs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq_19Jun03:*
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5115	98.3	7742	25	ABZ24689 Human cell growth, CDNA encoding mouse
2	4085.6	78.5	6395	20	AAV99308 Nucleotide sequence
3	4085.6	78.5	6395	22	AAV99308 Human CDNA sequence
4	2434.4	46.8	2550	22	AAH14878 Human ubiquitin re
5	1311	25.2	1635	24	ABQ75898 Human polynucleoti
6	1097.8	21.1	6840	22	AAK51709 Human polynucleoti
7	1088.4	20.9	6850	22	AAK52693 Human polynucleoti
8	996.2	19.1	1001	20	AAK03300 Partial CDNA encod

9	996.2	19.1	1001	22	AAK68934	Nucleotide sequenc
10	812	15.6	3327	20	AAK35730	CDNA encoding a pr
11	812	15.6	3502	20	AAK35731	CDNA encoding a pr
12	652.8	12.5	733	21	AAA02411	Human colon cancer
13	634.6	12.2	756	21	AAA02327	Human colon cancer
14	534.6	10.3	712	22	AAH07621	Human CDNA clone (
15	450.8	8.7	777	23	AAH70068	DNA encoding novel
16	429	8.2	1464	23	AAH70104	DNA encoding novel
17	373.2	7.2	1772	24	AAH90577	Human polynucleoti
18	271.4	5.2	1172	23	AAH70851	DNA encoding novel
19	267	5.1	2616	24	ABZ70161	Fibrinogen 9.57 co
20	249	4.8	972	24	AAK35394	Human CDNA encodin
21	246.8	4.7	807	20	AAK35891	Gastric cancer ass
22	238.4	4.6	264	21	ACA03886	Human colon cancer
23	218.2	4.2	4573	25	ACA03886	CDNA upregulated i
24	195	3.7	6278	23	ABH14799	Drosophila melanog
25	191.6	3.7	800	20	AAK39892	Gastric cancer ass
26	159	3.1	455	20	AAH89674	EST clone DA490.
27	142	2.7	643	24	ABH89546	Human polynucleoti
28	132	2.5	13134	23	ABH14798	Drosophila melanog
29	128	2.5	529	22	AAK62305	Human immune/haema
30	116.6	2.2	308	16	AAH21876	Human gene signatu
31	104.8	2.0	507	23	ABV59104	Human prostate exp
32	100.6	1.9	1382	25	ABH82975	Human lung specific
33	71	1.4	7428	22	ABH07164	Human pancreatic c
34	71	1.4	7428	25	ABH16907	Human secreted pro
35	71	1.4	7428	25	ABZ67490	Human secreted pro
36	66	1.3	8676	22	ABA07159	Human pancreatic c
37	66	1.3	31931	22	ABA07160	Human pancreatic c
38	66	1.3	51961	25	ABH16905	Human secreted pro
39	66	1.3	51961	25	ABZ67488	Human secreted pro
40	60	1.2	60	24	ABN36159	Plasmodium falcipa
41	54.8	1.1	2793	21	AAH070146	Chemically pretrea
42	49.4	0.9	12409	24	AAH63313	Human CDNA sequenc
43	49.2	0.9	1727	22	AAH17089	Human CDNA sequenc
44	48.6	0.9	4590	22	AAH24065	Yeast AOD9604-asso
45	46.2	0.9	581	22	AAH09232	Human CDNA clone (

ALIGNMENTS

RESULT 1	ABZ24689	ABZ24689 standard; CDNA: 7742 BP.
ID	ABZ24689	
XX	ABZ24689:	
XX	07-APR-2003 (first entry)	
DT	Human cell growth, differentiation and death protein CGDD-1 cDNA.	
XX		
DE	CGDD-1, cell growth, cell differentiation; cell death; human;	
KW	cytosolic; antiarteriosclerotic; hepatotropic; antiinflammatory;	
KW	antiproliferative; antineoplastic; ophthalmological; auditory;	
KW	anticonvulsant; cerebroprotective; nootropic; neuroprotective;	
KW	antiparkinsonian; neuroleptic; tranquilizer; immunosuppressive;	
KW	anti-HIV; antiallergic; antidiabetic; antidiabetic;	
KW	dermatological; nephrotoxic; antirheumatic; antiarthritic;	
KW	antitumor; vulnery; virucide; antibacterial; fungicide;	
KW	antiparasitic; protozoacide; antihelminthic; antifertility;	
KW	gynaecological; ubiquitin protein ligase; enzyme; gene therapy;	
KW	microarray; gene; ss.	
XX		
OS	Homo sapiens.	
XX		
FT	Key	Location/Qualifiers
FT	CDS	13..5229
FT		/*tag= a
FT		/product= "Human CGDD-1"
XX		
PN	W0200297032-A2.	

PD 05-DEC-2002.
XX 05-APR-2002; 2002WO-US11152.
XX 06-APR-2001; 2001US-282110P.
PR 11-APR-2001; 2001US-283294P.
PR 26-APR-2001; 2001US-286820P.
PR 27-APR-2001; 2001US-287228P.
PR 18-MAY-2001; 2001US-291662P.
PR 16-MAY-2001; 2001US-291846P.
PR 25-MAY-2001; 2001US-293727P.
PR 01-JUN-2001; 2001US-295263P.
PR 01-JUN-2001; 2001US-295340P.
PR 15-JAN-2002; 2002US-349705P.
XX (INCY-) INCYTE GENOMICS INC.
XX Azimzai Y, Au-Young JK, Batra S, Baughn MR, Becha SD, Borowsky ML;
PI Burford N, Ding L, Elliott VS, Emerling BM, Gandhi AR, Gietzen KJ;
PI Griffin JA, Hafalia AJA, Honchell CD, Lal PG, Lee SY, Lu DAM;
PI Arvizu CS, Ramkumar J, Reddy R, Sanjanwala MM, Tang YT, Walla NK;
PI Wang YE, Warren BA, Xu Y, Yang J, Yao MG, Yue H, Zebajadian Y;
XX WPI: 2003-140453/13.
DR P-PSDB; ABP58330.
XX
PT Novel human proteins associated with cell growth, differentiation and
PT death, useful for treating, diagnosing or preventing cancer,
PT developmental, neurological, reproductive or autoimmune/inflammatory
PT disorders
XX
PS Claim 5; Page 216-218; 238pp; English.
XX
CC The present sequence is that of Incyte clone 1351608CBI encoding
CC human CGDD-1, a novel protein associated with cell growth,
CC differentiation and death. A representative cDNA library for the
CC polynucleotide is PGANN01 from paraganglionic tumour tissue.
CC Structural features establish the encoded protein as being
CC associated with cell growth, differentiation and death, with
CC further evidence suggesting it to be a ubiquitin protein ligase.
CC The invention is based on novel human CGDD-1 to -21 proteins (see
CC ABP58330-50), the polynucleotides encoding them (see ABP24689-709),
CC and to the use of these for the diagnosis, treatment or prevention
CC of cell proliferative disorders including cancer, developmental
CC disorders, neurological disorders, autoimmune disorders,
CC reproductive disorders, and disorders of the placenta, and in the
CC assessment of the effects of exogenous compounds on the activity
CC and expression of proteins and nucleic acids associated with cell
CC growth, differentiation and death. CGDD polynucleotides are also
CC used in a claimed microarray and in a claimed method of generating
CC an expression profile of a sample.
XX
SQ Sequence 7742 BP; 2388 A; 1423 C; 1622 G; 2309 T; 0 other;
Query Match 98.3%; Score 5115; DB 25; Length 7742;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 5168; Conservative 10; Mismatches 6; Indels 45; Gaps 1;
QY 22 GGTACTGAGAGATGGAATACAGCGGGAGTTACCCAGACCCCTCAGCGCTCGGCATCT 81
DB 1 GGTACTGAGAGATGGAATACAGCGGGAGTTACCCAGACCCCTCAGCGCTCGGCATCT 60
QY 82 TGGTGGGATCAGCAAGTTCATTTTATCTGCTTTCTTCATCATCTTTGGCACAAATGGTG 141
DB 61 TGGTGGGATCAGCAAGTTCATTTTATCTGCTTTCTTCATCATCTTTGGCACAAATGGTG 120
QY 142 CCAGAAATTTACTTTGCTGAAATGGACCCAGACTTGGAAAAGCAGGAGGAAAGTGACAA 201
DB 121 CCAGAAATTTACTTTGCTGAAATGGACCCAGACTTGGAAAAGCAGGAGGAAAGTGACAA 180
QY 202 ATGTCAATATTCACCTCCACTGGATGGTACTTATTTGGAGAAGATCCAGATATTTGCTTA 261
DB 181 ATGTCAATATTCACCTCCACTGGATGGTACTTATTTGGAGAAGATCCAGATATTTGCTTA 240

QY 262 GAGAAATTGAAGCACAGTGGAGCAATTTCACTTTGTGGGAGGGTTTCAAAAGTGGAGAG 321
DB 241 GAGAAATTGAAGCACAGTGGAGCAATTTCACTTTGTGGGAGGGTTTCAAAAGTGGAGAG 300
QY 322 ACAACCTATCTTGCAGGGATTGTGCAATTTGATCCCAACATGTGTACTCTGTATGACTGC 381
DB 301 ACAACCTATCTTGCAGGGATTGTGCAATTTGATCCCAACATGTGTACTCTGTATGACTGC 360
QY 382 TTCCAGGACAGTGTTCATAAAATCATCGTTACAAGATGCATCTTCTACTGGAGGAGGG 441
DB 361 TTCCAGGACAGTGTTCATAAAATCATCGTTACAAGATGCATCTTCTACTGGAGGAGGG 420
QY 442 TTCTGTGACTGTGGAGACACAGAGGCATGGAACATGCGCCCTTTTGTGTAATCATGAA 501
DB 421 TTCTGTGACTGTGGAGACACAGAGGCATGGAACATGCGCCCTTTTGTGTAATCATGAA 480
QY 502 CTTGGAAGACAGTACTATATAAGAGAAATTCACGCTGTCGTTGTAATGAAGAGGTAAT 561
DB 481 CTTGGAAGACAGTACTATATAAGAGAAATTCACGCTGTCGTTGTAATGAAGAGGTAAT 540
QY 562 GTCCAAAGCCAGGAAATATTTCTTCACTGATATAATATGTCGTAGAAATGACTATATGG 621
DB 541 GTCCAAAGCCAGGAAATATTTCTTCACTGATATAATATGTCGTAGAAATGACTATATGG 600
QY 622 GAAGAGAAAGAACTGCTCTGAACTCCAGATAAGGKNR----- 663
DB 601 GAAGAGAAAGAACTGCTCTGAACTCCAGATAAGGAGGAGAAAAATGAAAGATACTAT 660
QY 664 -----YVCVNDHSHSYDHGTCTATATACAGCTCAACAAAGA 696
DB 661 TGTGCTCTTTCAATGATGAACACCATTCATATACCCAGCTCATATACAGCTCAACAAAGA 720
QY 697 GCTCTTGAAGTGTGAGCTGCGAGAGCCAGTTGATACCATCTGCCATTTGACAAAGAGGGT 756
DB 721 GCTCTTGAAGTGTGAGCTGCGAGAGCCAGTTGATACCATCTGCCATTTGACAAAGAGGGT 780
QY 757 CGTCGGGCTGTTAAAGCGGGAGCTTATGCTGCTTGCAGGAGCAAGCAAGGAGGATATAAG 816
DB 781 CGTCGGGCTGTTAAAGCGGGAGCTTATGCTGCTTGCAGGAGCAAGCAAGGAGGATATAAG 840
QY 817 AGTCATTCAGAAATGCTCTCAACATCCACTTCTAGAGATTTATACACTCAGAGATT 876
DB 841 AGTCATTCAGAAATGCTCTCAACATCCACTTCTAGAGATTTATACACTCAGAGATT 900
QY 877 ATGCTCATCAGAAATTTGCTTTGCTGCTTGGTTCCTGGATGAACAAATTTATGAGCTAT 936
DB 901 ATGCTCATCAGAAATTTGCTTTGCTGCTTGGTTCCTGGATGAACAAATTTATGAGCTAT 960
QY 937 TCAAGTGAATTTAGGAGATCTTTTGCAGGATGCTTTAGAGAGAACCTGACTCGGAG 996
DB 961 TCAAGTGAATTTAGGAGATCTTTTGCAGGATGCTTTAGAGAGAACCTGACTCGGAG 1020
QY 997 AATCCCTGCTCTAAGCAGGTTAATGCTTTGGGATGCAAGCTTTATAAAGTGCCTCGT 1056
DB 1021 AATCCCTGCTCTAAGCAGGTTAATGCTTTGGGATGCAAGCTTTATAAAGTGCCTCGT 1080
QY 1057 AAGATCCTTCATGAATTTGATCTTCAGGAGTTTATTTATGGAGATGGAATACAAAACCTC 1116
DB 1081 AAGATCCTTCATGAATTTGATCTTCAGGAGTTTATTTATGGAGATGGAATACAAAACCTC 1140
QY 1117 TTTGCTATGGAATTTGTGAAGTATTATAAACAACTGCAGAAAGAAATATATCAGTGATGAT 1176
DB 1141 TTTGCTATGGAATTTGTGAAGTATTATAAACAACTGCAGAAAGAAATATATCAGTGATGAT 1200
QY 1177 CATGACAGAGATATCTCTATAAATGCTGCTTTTCACTGATGTTTACTGTTCTTACTCTG 1236
DB 1201 CATGACAGAGATATCTCTATAAATGCTGCTTTTCACTGATGTTTACTGTTCTTACTCTG 1260
QY 1237 GCTCGACATCTTATGAAGACAGCAATGTTATCTGTCATTTACTGAAACTCTGCTAGAA 1296
DB 1261 GCTCGACATCTTATGAAGACAGCAATGTTATCTGTCATTTACTGAAACTCTGCTAGAA 1320

QY	1297	TTTTTACGTGACTTCTGGACAGACACATTAATTCAACTTCCAGGGTTATAGCCAGAC	1356
Db	1321	GTTTTACCTGGTACTCTGGACAGGAAACAATAAATTCACCTCCAGGGTTATAGCCAGAC	1380
QY	1357	AAATTGGGAAGATATATGACAGTAAATATGTACCTAAAGTATATCTCGATAGCAAAACC	1416
Db	1381	AAATTGGGAAGATATATGACAGTAAATATGTACCTAAAGTATATCTCGATAGCAAAACC	1440
QY	1417	ACAAATATGGACAGAAAGATTAAAGATGACAGTTCCTTGAAGGTTTTCGATCTTTTGGAG	1476
Db	1441	ACAAATATGGACAGAAAGATTAAAGATGACAGTTCCTTGAAGGTTTTCGATCTTTTGGAG	1500
QY	1477	ATTCTTACCTGTATGACAGGGAATGGAAAGAAATCCGAAGACAGTTGGGCAACACATTGAA	1536
Db	1501	ATTCTTACCTGTATGACAGGGAATGGAAAGAAATCCGAAGACAGTTGGGCAACACATTGAA	1560
QY	1537	GTGGAATCCTGATTGGGAGGCGCATTTGCTATACAGATGCAATGGAAGAAATATTTCCTC	1596
Db	1561	GTGGAATCCTGATTGGGAGGCGCATTTGCTATACAGATGCAATGGAAGAAATATTTCCTC	1620
QY	1597	ATGTTCCAGAGTGTGTGCTTTGATGAGAAGAACTTACTTGTGGCTTATAAAGAAATG	1656
Db	1621	ATGTTCCAGAGTGTGTGCTTTGATGAGAAGAACTTACTTGTGGCTTATAAAGAAATG	1680
QY	1657	CACAAAGCTGTGATGAGGTGCGACATGACAGTTTCAATCTAGTAGCAAGACAGTAGTACAA	1716
Db	1681	CACAAAGCTGTGATGAGGTGCGACATGACAGTTTCAATCTAGTAGCAAGACAGTAGTACAA	1740
QY	1717	TTCGTGTGACATATTTTGGAAACAAAGCCACAGACATCTGAGATCTTGTAGCATAT	1776
Db	1741	TTCGTGTGACATATTTTGGAAACAAAGCCACAGACATCTGAGATCTTGTAGCATAT	1800
QY	1777	CATCTGCGACATCTCTAGGACCCCTGTGCGTTCATGTAGCTTAAACAGGCTGGTGCT	1836
Db	1801	CATCTGCGACATCTCTAGGACCCCTGTGCGTTCATGTAGCTTAAACAGGCTGGTGCT	1860
QY	1837	GTTTCAAGACTGCAATGAATTTGTGCTTTTGAGACTTTCAAGTAGAGGTACTAGTGAA	1896
Db	1861	GTTTCAAGACTGCAATGAATTTGTGCTTTTGAGACTTTCAAGTAGAGGTACTAGTGAA	1920
QY	1897	TATCCTTAGTGTGTCGTGGTGTGGTGGCCAGGTTTGTGCTGAGATGTGGCCAAAGAAAT	1956
Db	1921	TATCCTTAGTGTGTCGTGGTGTGGTGGCCAGGTTTGTGCTGAGATGTGGCCAAAGAAAT	1980
QY	1957	GGAAGTCTCTTATTAAGCCAGAGTGTTTTATTACCAAGATGTTAAGTCAGAGAAAGAAATG	2016
Db	1981	GGAAGTCTCTTATTAAGCCAGAGTGTTTTATTACCAAGATGTTAAGTCAGAGAAAGAAATG	2040
QY	2017	TATGATTAAGAATATCATCATGCTTCAGATGTGGTCATCTTTAATGATCCCAATTAAGTTC	2076
Db	2041	TATGATTAAGAATATCATCATGCTTCAGATGTGGTCATCTTTAATGATCCCAATTAAGTTC	2100
QY	2077	TTGTCTACTGCTACTTCAGAGGTATGAACTTGGCCGAGGCTTTTAAACAAGACCATATCTACA	2136
Db	2101	TTGTCTACTGCTACTTCAGAGGTATGAACTTGGCCGAGGCTTTTAAACAAGACCATATCTACA	2160
QY	2137	AAAGACAGCAGATTGATTAAGAAACAATATATATACATATATAGAGAAATGCTTCAGTCTC	2196
Db	2161	AAAGACAGCAGATTGATTAAGAAACAATATATATACATATATAGAGAAATGCTTCAGTCTC	2220
QY	2197	ATCTATATTTGTGGGTGAGCCTTATGACTCTGAGGTGGAAATGTGACCAAGAAAGAGGTC	2256
Db	2221	ATCTATATTTGTGGGTGAGCCTTATGACTCTGAGGTGGAAATGTGACCAAGAAAGAGGTC	2280
QY	2257	ACATATGAGAAATCATCTACCTTGCTTGCATTTGAATACCCATGCCACACAGTGGCATTTGCC	2316
Db	2281	ACATATGAGAAATCATCTACCTTGCTTGCATTTGAATACCCATGCCACACAGTGGCATTTGCC	2340
QY	2317	AAAAATTTTACCTGGAATGAAAAATATGAAACGTGGCTTAGAGAAATGTCATATAAACAAAGTG	2376
Db	2341	AAAAATTTTACCTGGAATGAAAAATATGAAACGTGGCTTAGAGAAATGTCATATAAACAAAGTG	2400
QY	2377	GCACACATTTTAAAGAAACCCAGGTGTATGACGGCCATGGAATTTATGAACTTAAAAAGATGAATCA	2436

Db	2401	GCACATTTAAGAAAACGAGGTGTATCAGGCCATGGAGTTTATGAACCTAAAAGATGAATCA	2460
OY	2437	CTGAAGAAGCTCAATATGTACTTTTATATTAATCTCAAAACCCAGCATAGAAAGGCTCAA	2496
Db	2451	CTGAAGAAGCTCAATATGTACTTTTATATTAATCTCAAAACCCAGCATAGAAAGGCTCAA	2520
OY	2497	CATATGCAGAAGAAAAGAGAAAACACAGAAAACAAAGATGAAAGCATTCGCCGCACACCA	2556
Db	2521	CATATGCAGAAGAAAAGAGAAAACACAGAAAACAAAGATGAAAGCATTCGCCGCACACCA	2588
OY	2557	CCTCTGAATTCCTGCCCTGCTTTCAGCAAAGTATTAACTTCTCACTGTGATATCATG	2618
Db	2581	CCTCCTGAATTCCTGCCCTGCTTTCAGCAAAGTATTAACTTCTCACTGTGATATCATG	2640
OY	2617	ATGTACATTCACGAGACCGTATTGAGGGGCAATAGACACAGATTCATACCTGGAGAC	2676
Db	2641	ATGTACATTCACGAGACCGTATTGAGGGGCAATAGACACAGATTCATACCTGGAGAC	2700
OY	2677	GAAGGAGTGTCCAAATGGCTTTTCATATTCGGCATGGGGTTTACTAGAGAGACAA	2736
Db	2701	GAAGGAGTGTCCAAATGGCTTTTCATATTCGGCATGGGGTTTACTAGAGAGACAA	2760
OY	2737	CAGCTTCAAAAAGCTCCTGAAGAAAGATTAACATTTGACTTTTATCATTAAGCTTCAAGA	2796
Db	2761	CAGCTTCAAAAAGCTCCTGAAGAAAGATTAACATTTGACTTTTATCATTAAGGCTTCAAGA	2820
OY	2797	TTGGGAATTCAGGCATGTAATTAATCAATTCGTTTTGGAAAAATCAAAAGAAATCCCGAG	2856
Db	2821	TTGGGAATTCAGGCATGTAATTAATCAATTCGTTTTGGAAAAATCAAAAGAAATCCCGAG	2880
OY	2857	TTAGAGGCCCAAGAGGACATGATTAACGTGGATCTTCAGATGTTTGAACACGTAAAGCA	2916
Db	2881	TTAGAGGCCCAAGAGGACATGATTAACGTGGATCTTCAGATGTTTGAACACGTAAAGCA	2940
OY	2917	TTAAGAGAAAATCTGTTTAAATTGTAGCAACACATCAGATCGGAATCTATTATAGAT	2976
Db	2941	TTAAGAGAAAATCTGTTTAAATTGTAGCAACACATCAGATCGGAATCTATTATAGAT	3000
OY	2977	GATGAGATTAATCTCAATGAATGAAGAAAACCAACAGAAAACAAAAGCTGCTGATGG	3036
Db	3001	GATGAGATTAATCTCAATGAATGAAGAAAACCAACAGAAAACAAAAGCTGCTGATGG	3060
OY	3037	CTACATGCCCAAGAGATCATGGCTCAGATGCTCCCTTACAGAAAACCTTATTTGAACCT	3096
Db	3061	CTACATGCCCAAGAGATCATGGCTCAGATGCTCCCTTACAGAAAACCTTATTTGAACCT	3120
OY	3097	CATAAATCATATGATGACATATACATCAGAAAATGCTGGGAAAGAGATTCATTATGAG	3156
Db	3121	CATAAATCATATGATGACATATACATCAGAAAATGCTGGGAAAGAGATTCATTATGAG	3180
OY	3157	GAAAGAGACACCCAGACGATCAGTGACACTCTGAANTGCTTTGGGCTCTAAACGGGCT	3216
Db	3181	GAAAGAGACACCCAGACGATCAGTGACACTCTGAANTGCTTTGGGCTCTAAACGGGCT	3240
OY	3217	CCATCTGTACTGAAAAAGAGGCTCTGACGTGCATCTCTTTCCCAAGAAAGACAGAGGCT	3276
Db	3241	CCATCTGTACTGAAAAAGAGGCTCTGACGTGCATCTCTTTCCCAAGAAAGACAGAGGCT	3300
OY	3277	AAAAATGAGAAATTAATGCCATGGTATTATCGGCTGTGTCCAGAAATCTACTGCTTAAAC	3336
Db	3301	AAAAATGAGAAATTAATGCCATGGTATTATCGGCTGTGTCCAGAAATCTACTGCTTAAAC	3360
OY	3337	CAGACAGAGGGGAAAACCCATGAGACTCTCAGAGAAAGCCCTAGAACCCACTTTATGATAT	3396
Db	3361	CAGACAGAGGGGAAAACCCATGAGACTCTCAGAGAAAGCCCTAGAACCCACTTTATGATAT	3420
OY	3397	CCACACTTTGGCATATGGAATTTATACAGAGAGCTGTGGCTCATGTAATGACAGCAGTGTGC	3456
Db	3421	CCACACTTTGGCATATGGAATTTATACAGAGAGCTGTGGCTCATGTAATGACAGCAGTGTGC	3480
OY	3457	TGGCAGAAAGTATTTTGAAGCTGTACAGCTGAAGCTCTCAGACGCCATTCATGTTGACCTT	3516

3481 TGCGAGAAGTATTTTGAAGCTGTACAGCTCAGCTCTCAGCAGCGCATTCATGTTGACCTT 3540
3517 TTTGACTTGAAAGTGGAGAAATATCTTGGCCCTCTTTGGAAATCTCTGTGCAATACTGTG 3576
3541 TTTGACTTGAAAGTGGAGAAATATCTTGGCCCTCTTTGGAAATCTCTGTGCAATACTGTG 3600
3577 ATCCCATATATCCCTTTGCAACCTCAAAAGATAAACAGTGAAGATCGAGATGCTCTTGCT 3636
3601 ATCCCATATATCCCTTTGCAACCTCAAAAGATAAACAGTGAAGATCGAGATGCTCTTGCT 3660
3637 CAACCTTTGACCCCTGGCAGGTGGATACAGACTGTTCTTGCCCAAGATATACAGGTTATAAT 3696
3661 CAACCTTTGACCCCTGGCAGGTGGATACAGACTGTTCTTGCCCAAGATATACAGGTTATAAT 3720
3697 ATAAGACATGCTAAAGAGGAAACCCAAATCCCTATTTTCTTTAATCAAGGAATGGAGAT 3756
3721 ATAAGACATGCTAAAGAGGAAACCCAAATCCCTATTTTCTTTAATCAAGGAATGGAGAT 3780
3757 TCTACTTTGGAGTTCCATTCCTCCAGTTTGGGCTTGAGTCTTCGATTAATATTCGA 3816
3781 TCTACTTTGGAGTTCCATTCCTCCAGTTTGGGCTTGAGTCTTCGATTAATATTCGA 3840
3817 ATAGCATCAAGGAATGGTTATCTCTTGCCCAACAAATTTATAGAATTGGAATGAAA 3876
3841 ATAGCATCAAGGAATGGTTATCTCTTGCCCAACAAATTTATAGAATTGGAATGAAA 3900
3877 GTGCCACCTGATGAAGGGATCCCTGAGTCCCATCTGACCTGGAGCACCTCGCTTTC 3936
3901 GTGCCACCTGATGAAGGGATCCCTGAGTCCCATCTGACCTGGAGCACCTCGCTTTC 3960
3937 ACTATCCAGCAATGAAATCTATTTGGGAGATGAAGGAAACCTCTGTTTGAGCACCT 3996
3961 ACTATCCAGCAATGAAATCTATTTGGGAGATGAAGGAAACCTCTGTTTGAGCACCT 4020
3997 CAAATAGCAGCATATGCTCTGAAGCATTAATGCAGTTTCAGTTGCACAGAGGATT 4056
4021 CAAATAGCAGCATATGCTCTGAAGCATTAATGCAGTTTCAGTTGCACAGAGGATT 4080
4057 ACCTGCTCCTCAGTCTGTATACAGAAACATCTGTTGCTCTCTATCAGTTGTTCTTCCT 4116
4081 ACCTGCTCCTCAGTCTGTATACAGAAACATCTGTTGCTCTCTATCAGTTGTTCTTCCT 4140
4117 AACATAAATCAGAAGATACACCATGCTCTCTCTATAGATCTGTTTATGTTTGGTG 4176
4141 AACATAAATCAGAAGATACACCATGCTCTCTCTATAGATCTGTTTATGTTTGGTG 4200
4177 GGTGCTGTGTAGCATTCCTCCATCCTGTGATTTGGGATGACCTCTGTATCTGCAGCCTTCT 4236
4201 GGTGCTGTGTAGCATTCCTCCATCCTGTGATTTGGGATGACCTCTGTATCTGCAGCCTTCT 4260
4237 TCAGTTAGTCTTCTCCATACCCACCTTTATCTCTCCATTTGATCACCATTGGCACACATG 4296
4261 TCAGTTAGTCTTCTCCATACCCACCTTTATCTCTCCATTTGATCACCATTGGCACACATG 4320
4297 CTTTCAGATACTACTACAGTAGACAGAGGCTACCCCTGCTCAGGTTCAAGAGACAGT 4356
4321 CTTTCAGATACTACTACAGTAGACAGAGGCTACCCCTGCTCAGGTTCAAGAGACAGT 4380
4357 GAAGAGGCTCATTCGCGATCTCTCTTCTTGCAAGAAATTTCTCAATATACAAGTGGCTCC 4416
4381 GAAGAGGCTCATTCGCGATCTCTCTTCTTGCAAGAAATTTCTCAATATACAAGTGGCTCC 4440
4417 ATTGGGTGTGATATTCCTGGCTGGTATTTGTTGGGTCTCACTGAAGATGGCATCACCCCT 4476
4441 ATTGGGTGTGATATTCCTGGCTGGTATTTGTTGGGTCTCACTGAAGATGGCATCACCCCT 4500
4477 TATCTTCGCTGTGCTGCATTTGTTTTCACATATTTACTTGGGGTAACCTCGCCTGAGGAA 4536
4501 TATCTTCGCTGTGCTGCATTTGTTTTCACATATTTACTTGGGGTAACCTCGCCTGAGGAA 4560
4537 CTGCATACCAATCTCCAGAGGAGATGACAGTGCACCTCTGTAGCTATCTATCTTTACCT 4596
4561 CTGCATACCAATCTCCAGAGGAGATGACAGTGCACCTCTGTAGCTATCTATCTTTACCT 4620

QY 4597 ACAAAATTTGTTCTGCTCTTCCAGGAATATTTGGGATACTGTAAGGCCCTTGTCTCAGAGG 4656
DDB 4621 ACAAAATTTGTTCTGCTCTTCCAGGAATATTTGGGATACTGTAAGGCCCTTGTCTCAGAGG 4680
QY 4657 CGGTGTCAGATCTGCTCTTAACTGTTTGAAGCAAAACACCGTGGTCAAGTAC 4716
DDB 4681 TGGTGTGAGATCTGCTCTTAACTGTTTGAAGCAAAACACCGTGGTCAAGTAC 4740
QY 4717 CCTAGAAAAGAAATAGTTTATAGAGCTTCTCTGATGACTATAGTCTCTCTGAATCAA 4776
DDB 4741 CCTAGAAAAGAAATAGTTTATAGAGCTTCTCTGATGACTATAGTCTCTCTGAATCAA 4800
QY 4777 GCTTCTCATTTACAGTGCCCGCTGTCAGATGATGAGCGAAAGCATCTCTCTCTGC 4836
DDB 4801 GCTTCTCATTTACAGTGCCCGCTGTCAGATGATGAGCGAAAGCATCTCTCTCTGC 4860
QY 4837 CTTTTCTGTGGGCTATACTATGTTCTCAGAACATTTTCTGCCAGAAATTTGTGAACGG 4896
DDB 4861 CTTTTCTGTGGGCTATACTATGTTCTCAGAACATTTTCTGCCAGAAATTTGTGAACGG 4920
QY 4897 GAAGAGTTGGAGCTTGCAATTTTTCACGCACTTCACTGTGGAGCGGAGTCTGCAATTTTC 4956
DDB 4921 GAAGAGTTGGAGCTTGCAATTTTTCACGCACTTCACTGTGGAGCGGAGTCTGCAATTTTC 4980
QY 4957 CTAAAAATCAGAGAATGCCGAGTGGTCTGTTGAAGTAAAGCCAGAGGCTGTGCCCTAT 5016
DDB 4981 CTAAAAATCAGAGAATGCCGAGTGGTCTGTTGAAGTAAAGCCAGAGGCTGTGCCCTAT 5040
QY 5017 CCAGCTCCTTACTTGGATGATATGGAGAAACAGACCTTGCCCTGAAGAGGGGCAACCCC 5076
DDB 5041 CCAGCTCCTTACTTGGATGATATGGAGAAACAGACCTTGCCCTGAAGAGGGGCAACCCC 5100
QY 5077 CTTTCAATTTATCTCTGAGCGGTATCGGAAGCTCCATTTGGTCTGGCAACAACACTGCAATT 5136
DDB 5101 CTTTCAATTTATCTCTGAGCGGTATCGGAAGCTCCATTTGGTCTGGCAACAACACTGCAATT 5160
QY 5137 ATAGAAGAGATTTGTAGAGCCAAAGAGACTAATCAGATGTTTATTTGGATTCACTGGCAG 5196
DDB 5161 ATAGAAGAGATTTGTAGAGCCAAAGAGACTAATCAGATGTTTATTTGGATTCACTGGCAG 5220
QY 5197 TTACTGTGA 5205
DDB 5221 TTACTGTGA 5229
RESULT 2
ID AAV99308 standard; cdna; 6395 BP.
XX AAV99308;
XX AC AAV99308;
XX XX 25-MAR-1999 (first entry)
XX DE cdna encoding mouse a ubiquitin-protein ligase, Ubql.
XX KW Ubiquitin-protein ligase; Ubql; mouse; ubiquitinylation; degradation;
XX KW N-end rule pathway; stress-related muscle wasting; inhibitor; screen; ss.
XX OS Mus sp.
XX FH Key Location/Qualifiers
XX FT CDS 115..5388
XX FT /*tag= a
XX FT /product= Ubql
XX XX US5861312-A.
XX XX 19-JAN-1999.
XX XX 02-DEC-1997; 97US-0982956.
XX XX 02-DEC-1997; 97US-0982956.

Dp	4075	GTGGCCATGATGACCTGGAGACGCTGTCGCTTCCACCATTCAGGCAATCGAAACCTGTGG	4134
OY	3964	GGAGATGAAGAAACCTCTGTTTGGAGCACTTAAATATAGCAGCATTAATGGTCTAAA	4023
Dp	4135	GGAGATGAAGAAACCTCATTTTGGAGCACTTCAAAATATAGCAGCATTAAGGCTGTGAAG	4194
OY	4024	GCATTATGAGAGTTTGGATGTGCACAGAGATTCACGTGCTCCAGATCGATTCAGAAA	4083
Dp	4195	GCAGTATGCAATTTGAGTTGGACACAGAGGCTACCTGCTCCAGGCTCTGATTCACAAA	4254
OY	4084	CATCTGGTGCCTCTTATCAAGTTGTTCTTCTTAACATTAATAATCAGAAATACACATGC	4143
Dp	4255	CATCTGCTCGGGCTCCGTGCAATTAATCTTCCATACGCAATCAGAAAATACACAGGC	4314
OY	4144	CTTCTGCTATAGATCTGTTCAATGTTTGGTGGGTCTGTTAGATTCOCATCTCTGG	4203
Dp	4315	CTTCTGCTGATGATCTTCCATGTTCTGGTGGGCGCAGCTTAAGGTGCCCAATCTTG	4374
OY	4204	TATTTGGATACACCTGTTGATATGTCACACCTCTCTCAATTAATGTTTCTTAACACCTT	4263
Dp	4375	TATTTGGATAGACACCGTGGATCTGCACACCGCTCCGCACTTAGTCTTATATTAACACCTC	4434
OY	4264	TATCTCTTCATTTGATACACCATGCGACACATGCTTCAGATCTACTTACAGTAGACA -	4321
Dp	4435	TATCTCTTCATCTGATACACCATGCGCAGCATGCTTCAGATCTCTTACACAGATACA	4494
OY	4322	-----CAGGCCATCCCTTGGCTCAGTTTCAABAAGACATGGAAGAGCTCATATCC	4371
Dp	4495	GATCTGTCTCAGAGGCGCGCGCTGAGAGGTGAAGAGTATGTAGAGAGCGCTGT	4554
OY	4372	GCATCTCTCTTCTTGGAGAAATTTCTCAATATCAAGTGGCTCATTTGGGTGTGATAT	4431
Dp	4555	GCATCTCTCTTCTTGGAGAGTGTCCAGACACAGACGCGCTCATCTGGTGGGTGTCT	4614
OY	4432	CTTGCGTGTATTTTGGGGTCTCACTAGAAATGGCATCAACCCCTTAATCTTGCTGTCT	4491
Dp	4615	CCGGGCTGGTACCTGTGGCTCTCCCTGAGGAAGGCAACCCCTTAATCTTGCTGTCT	4674
OY	4492	GCATTTGTTTTCCACTATTTACTTGGGGTAACTCCGCTCAGAGAACTGCATACCAATTC	4551
Dp	4675	GCATCTGTTTCCACTATTTACTTGGGGTAACTCCGCTGAGAAAGCTTTGGCCAAATTC	4734
OY	4552	GCAGAAAGAGATGACACGCTGAGCATATCTTACCTACCAAAATTTCTCTCG	4611
Dp	4735	GCTGAAGAGATTCAGTGCACCTGTGAGCTATATCTTAAACCAAAATTTCTCTCG	4794
OY	4612	CTCTTCCAGGAATTTGGATACTGTAAAGCCCTTGGCTCCAGAGCGGTGTGCAGATCT	4671
Dp	4795	CTTTTCCAGGAATTTGGATACCATTAAGGCCCTTACTACAGAGTGGTGTGGAGATCT	4854
OY	4672	GCTTACTTAACTGTTTGAAGCAAAAACACGCTGGTCAGTAGACCTAGAAAAAGAAAT	4731
Dp	4855	GCTTACTTCAAGTCTTGAAGCAAAAAGTCTGTGGTCAGGTACCTTAGAAAAAGAAAT	4914
OY	4732	AGTTTGAATAGAGCTTCTGATGATATAGTGCCTCTGATTCAGAGCTTTCATTTCAGG	4791
Dp	4915	AGTTTGAATAGAGCTTCTGAGAGCTACAGCTCTTCTTAATTCAGGCTTCCACTTATAGG	4974
OY	4792	TGCCACAGGCTGCGAGATGATGACCGAAAGCATCTGCTCCCTTTTGTGGGGCT	4851
Dp	4975	TGTCACAGGCTGCGAGATGATGACCGAAAGCATCTGCTCTGCTCTTTTGTGGGGCC	5034
OY	4852	ATACTATGTTCTCGAACAATTTGCTGCCAGGAAATGTGAACGGGGAAGAGGTTGAGCT	4911
Dp	5035	ATCTGTGTTCTCGAACAATCTGTGGCCAAAGAAATAGTAATGGGGAAGAGGTTGAGCG	5094
OY	4912	TGCATTTTTCACGCACTTCACTGTGGAAGCGGAGCTGTGATTTTCTTAATAATCAGAGAA	4971
Dp	5095	TGCTTTTTCATGCGCTTCATTTGTGGTCTGGAATCTGCATTTTCTTAATAATCAGAGAA	5154
OY	4972	TGCCAGATGCTGCTGGTGAAGGTAAGGACGAGAGGCTGTGCTATTCACAGCTCTTACTGT	5033
Dp	5155	TGCAGGATGCTGCTGGTGAAGGAAAGCCAGAGAGCTGTGCTATTCACAGCTCTTACTGT	5214

QY	5032	GATGAATATGGAGAAACAGACCCTGGCCCTGGAAGAGGGCAACCCCTTCATTATACGT	5091
Db	5215	GATGAATATGGAGAAACAGACCCTGAGGGCTTAAGAGAGAAACCCATTCATTATCTGGG	5274
QY	5092	GAGCGGTATCGGAGAGCTCATTTTGGTGGCAACACCTGCATTATAGAGAGATTGCT	5155
Db	5275	GAGCGGTATCGGAGAGCTGCATTTTGGTCTGGCAACGCACTGCATTATAGGAAGATTGCT	5338
QY	5152	AGGAGCCACAGCACTAATCAGATTCTTATTTGGATTCACTGGCAGTTAAGTGA	5205
Db	5335	CGGAGCCAGGAGCACTAATCAGATGCTATTTTGGATTTAAGTCACTTCACTGCA	5388

RESULT 3	
AAC86933	
ID AAC86933	standard; cDNA; 6395 BP.

DT 02-APR-2001 (first entry)

Nucleotide sequence of Mouse Ubr1 protein.

KM Ubir1; E3-type protein; ubiquitin system; ubiquitin-protein ligase;
 KM N-end rule pathway; intracellular pathogen; *Lysteria monocytogenes*
 KM *Yersinia enterocolitica*; muscle wasting; infection; ss.

Mus sp.

FH	Key	Location/Qualifiers
FT	CDS	115..5388
FT		/*tag= a
FT		/product= "ubrl"

PN US6159732-A.

PD 12-DEC-2000.

PF 11-JAN-1999; 99US-0228317.

PR 02-DEC-1997; 97US-0982956.

PA (CALY) CALIFORNIA INST OF TECHNOLOGY

PI Kwon YT, Varshavsky A;

DR WPI; 2001-090278/10.

PT Inhibiting the N-end rule pathway in mammalian cells for treating
PT infections and various diseases associated with muscle tissue wasting,
PT by inhibiting the expression of Ubr1 gene -
XX
PS Claim 3; Column 15-26; 18pp; English.

Claim 3; Column 15-28; 18pp; English.

The present sequence encodes a murine ubl1 enzyme. Ubl1 is an E3-type protein of the ubiquitin system. Specifically, it is a ubiquitin-protein ligase. The enzyme is specific for destabilizing residues exposed at the N-terminus of protein substrates. Inhibition of the expression of ubl1 gene in a cell results in inhibition of the N-end rule pathway. The method is used for treatment of mammalian cells infected with an intracellular pathogen, e.g. *Yersinia monocytogenes* or *Yersinia enterocolitica*. Inhibition of N-end rule pathway is also useful for treating various diseases associated with wasting of muscle tissue and infections.

SQ Sequence 6395 BP; 1802 A; 1376 C; 1519 G; 1698 T; 0 other;

Query Match	78.5%	Score 4085.6	DB 22	Length 6395
Best Local Similarity	86.6%	Pred. No. 0		
Matches 4367; Conservative	9	Mismatches 629	Indels 69	Gaps 4

Qy	1	ATGCCGACGAGGAGGCTGGAGGTACTGAGAGATGGAATACAGCCGGAGTTACCCCGAG	60
Db	115	ATGCGGACGAAGAGATGGAGCGCGCGAGGATGGACGCTCAGCCCGAGAGCCTCCCGT	174
Qy	61	ACCCTCAGCGTCTGGCATCTCGTGGGATCAGCAAGTTGATTTTATACGTCTTCTTG	120
Db	175	GCCCGCAGCGCGCGCATCGTGGTGGGATCAGCAAGTTGATTTCTATACTGTTCTT	234
Qy	121	CATCATTTGGCCACAATTTGGTCCAGAAATTTACTTTGCTGAAATGGACCCAGACTTGGAA	180
Db	235	CATCATTTGGCCACAATTTAGTGGCCAGAAATTTATTTGCTGAGATGGACCCAGATTTGGAA	294
Qy	181	AAGCAGAGGAAAGTGTACAAATGTCAATATTCACTCCACTGGAAATGGTACTATTGGGA	240
Db	295	AAGCAAGAAGAGAGTGTACAGATGTCAATACTCACTCCTTTGGAGTGGTACTATTGGGA	354
Qy	241	GAAGATCCAGATATTTCCTTTAGAGAAATTCGACACAGTGGAGCATTTACGTTTGGGG	300
Db	355	GAGATCCGGATATTTCCTTTAGAGAAATTTAAACACACAGTGGAGCGTTCCAGTTGTGGG	414
Qy	301	AGGTTTTTCAAAAGTGGAGACAACTATTCTTTGACGGGATGTGCAATTGATCCCAACA	360
Db	415	AGGTTTTTCAAAAGTGGAGAAACAACATATCTCTGTAGGATTTGTGCAATTGATCCCAAG	474
Qy	361	TGTGTACTGTATGGACTCTTCCAGGACAGTGTTCATAAAATCATCTTTACAAGATG	420
Db	475	TGTGTCTCTGTATGGACTCTTCCAAAGTAGTGTTCATAAAACCATCCTTACAAGATG	534
Qy	421	CATACTTCTACTGGAGGAGGGTCTGTGACTGTGGAGACACAGAGGCATGGAAAATGGC	480
Db	535	CATACTTCTACTGGAGGAGGGCTTCTGTGACTGTGGAGACACAGAAGCGTGGAAAATGGC	594
Qy	481	CTTTTTTGTGTAAATCATGAACCTGGAAAGACAGGTACTATAAAGAGAAATTCACGCTGT	540
Db	595	CTTTTTTGTGTGATCAGAGCTGGAAAGACAGGTACTATAAAGAGAGACTTACATTCG	654
Qy	541	CCGTTGAATGAAGAGTAAATTTGTCCAAAGCCAGGAAAATATTCTTTCAGTGTAAATAT	600
Db	655	CCATTGAATGAAGAGTGAATTTGTCTCAAGCCAGGAGAAATATTCCTTCGGTGTAAATATC	714
Qy	601	GTCTAGAAATGACTATATGGGAGAGGAAAAGAACTGCCTCCTGAACTCCAGATGAAGG	660
Db	715	ATTGTAGAAATGACTATATGGGAAGAAGAAAGAAATTCCTCCTGAACTCCAGATGAAGG	774
Qy	661	KNR-----YVCVNDHHSDH	675
Db	775	GAGAAAAATGACGATACTATTGTGCTCTTTCAACGATGAGCACCATTGCTATGATCAT	834
Qy	676	GTCAATACAGCCTTACAAGAGCTCTTGACTGTGAGCTCGCAGAGGCCAGTTGCAATACC	735
Db	835	GTGATCTACAGCTGCGAGAGAGCTCTAGATTGGAGCTTCGAGAGGCACAGCTGCACAG	894
Qy	736	ACTGCCATTGACAAAGAGGTGCTCGGGCTGTTAAGCGGGAGCTTAATGCTGTGGCCAG	795
Db	895	ACTGCCATCGACAAAGAGGTGCGCGGCTGTCAAAGCAGGTGTGTATGCCACTTGGCCAG	954
Qy	796	GAAGCAAGGAGATATAAGAGTCAATTCAGAAAAATGCTCTCAACATCCACTTCATGTA	855
Db	955	GAAGCAAGGAGGATATAAGAGTCAATTCAGAAAAATGCTCTCAACATCCACTTCATGTA	1014
Qy	856	GAAGTATTACACTCAGAGATTATGGCTCAATCAGAAAAATTCGTTTGGCTCTTGGTTCCTGG	915
Db	1015	GAAGTGTGCATCCGTGGTTATGGCTCACCAGAAATTCGCTCTGCGCTTGGCTCTCTGG	1074
Qy	916	ATGAACAAAATATGAGCTATTCAGTGTACTTTAGGCAGATCTTTTGGCCAAAGCATGCC	975
Db	1075	ATGAACAAAATATGAGCTATTCAGTGTACTTTAGACAGATATTTTGGCCAGGCTGCTCT	1134
Qy	976	AGAGAAGAACTGACTCGGAGATCCCTGTCTCATAGCAGGTTAATGCTTTGGGATGCA	1035
Db	1135	GTAGAAGAACTGCTCTGAAAATCCCTGTCTTATAGCAGAGCTAATGCTTTGGGATGCA	1194
Qy	1036	AAGCTTTTAAAGGTGCCCTTAAGATCCCTTCATGAATTTGATCTTCACAGCTTTTTTATG	1095

Db 2275 TTTACAAGACCATATCCACAAGAACGAGATTGATTAAACGATATAATACATTATA 2334
QY 2176 GAAGAAATGCTTCAGAGCTCTCATCTATATTTGGGTGAGCGTTATGTACCTGGAGTGGGA 2235
Db 2335 GAAGAAATGCTTCAGAGCTCTCATCTATATTTGGGTGAGCGTTATGTACCTGGAGTGGGA 2394
QY 2236 AATGACCAAGAGAGAGTGCATATGAGAGAAATCATCTGCTTGGTTCATTTGAACCC 2295
Db 2395 AATGTTACAGAGAGAGTGTATATGAGAGAGATTACTACCTTACTTTGCAATGAGCCC 2454
QY 2296 ATGCCACAAGTGCATTGCCAAAAATTTACCTGGAATGAAATATGAAACTGGCTTA 2355
Db 2445 ATGCCACAAGTGCATTGCCAAAAATTTACCTGGAATGAAATATGAAACTGGCTTA 2514
QY 2356 GAGATGTCATTAACAAAGTGGCCCATTTAAGAAACGAGTGTATACAGGCCATGAGATT 2415
Db 2515 GAGATGTCATTAACAAAGTGGCCCATTTAAGAAACGAGTGTGTGGGCCATGAGATT 2574
QY 2416 TATGAACATAAAGATGATCACTGAAGACTCAATATGTACTTTATCATACCTCAAA 2475
Db 2575 TATGAATTTGAAGATGAATTCATCTGAAGACTTCAATATGTACTTTATCATATTTCTAAA 2634
QY 2476 ACCGACATAGCAAGGCTGAACATATGCAAGAAAGAGAAAAACAAGAAACAAGAT 2535
Db 2635 ACACAGCATAGCAAGGCTGAACATATGCAAGAAAGAGAAAAACAAGAAATAAGAT 2694
QY 2356 GAAGCATTTGCCGCCACCAACCTCTCTTAATTTGCCCTGCTTTCAGCAAGTGAATAC 2595
Db 2695 GAAGCATTTGCCGCCGCCACCTCTCTCAAGATTTGCCCTGCTTTCAGCAAGTGAATAC 2754
QY 2596 CTTCGACATGATGATATGATATGATATGATATGATATGATATGATATGATATGATATG 2655
Db 2755 CTGCTCAGCTGTGATGTATGATATGATATGATATGATATGATATGATATGATATGAT 2814
QY 2656 ACAGATTTCACTTGTGAGACGAGAGATGCTCCAAATGGCTTTTCAATTTTCGCATG 2715
Db 2815 ACGAGTGTATATCTGTGAGAGAGAGATGCTGCAAGATGCGCTTCATATATTTGCACATG 2874
QY 2716 GGTTTACTAGAGAGAGACCAACAGCTTCAAAAAAGCTCTGGAAGAAAGTAACATTGAC 2775
Db 2875 GGTGTGCTGAGAGAGAGACAGAGCTTCAGAAAGCTCCGAGAGAAAGTGGCTTTTGAC 2934
QY 2776 TTTTATCTAAGGCTTCAAGATTTGGGAACTTCAGCAT-----GAATATACAAAG 2826
Db 2935 TTTTACCTTAAGCTTCAAGATTTGGGAACTTCAGCATGAATGCTCAGAAATTAACAAAG 2994
QY 2827 CTTTTGGAAAACTCAAGGAATTCGCCAGTTAGAAGCCAGACAGACATGATTAACGTTG 2886
Db 2995 CTCTTGGAAAACTCAAGGAATTCGCCCAATTAGAGGCCAGAGACATGATTAACATGG 3054
QY 2887 ATACTTCAGATGTTTGACACAGTGAAGGATTTAGAGAAAAATCTTGTTAATTTGATGA 2946
Db 3055 ATACTTCAGATGTTTGACACAGTGAAGGATTTAGAGAAAAATCTTGTTAATTTGATGA 3114
QY 2947 ACCCATCAGAGTGGGATCTTATTAAGATGATGATTTACTCATGATTAAGAAAAAGCA 3006
Db 3115 ACCCATCAGAGTGGGATCTTATTAAGATGATGATTTACTCATGATTAAGAAAAAGCA 3174
QY 3007 GAAGAGAAAAAGAAAGCTGTAAGCTGTAGGCTACATGCGCAAGAAATCATGCTGATGAG 3066
Db 3175 GAAGAGAAAGAAAGCTGTAAGCTGTAGGCTTCAATGCCAGAAATCATGCTGCCAGATG 3234
QY 3067 TCTGCTTACAGAAAACTTCATTTGAATCTATAACTCATGTATGACAAATCATACAGAA 3126
Db 3235 TCTGCTTACAGAAAACTTCATTTGAATCTATAACTCATGTATGATATGATATGATATG 3294
QY 3127 ATGCTGGGAAAGAGATTTCAATTAAGAGAGAGACACCCACAGCTGATGATCACTAC 3186
Db 3295 GTAAAGAGGAAAGAGACTTCATTAATGAGAGAGAGACACTGACAGTGAAGGCGC 3354
QY 3187 TCTGAATTTGCTTTGGGTCTTAACAGGAGGCTCATGTTTACTGAAAAGAGAGGCTGACG 3246
Db 3355 TCTGAATTTGCTTTGGGTCTTAACAGGAGGCGCTGTTTACCGAAAGAGAGGCTGCTGACG 3414

QY 3247 TGCATCTTTGCCAAGAGAACAGAGAGTGAATAATAGAAAAATATGCGATGTATATCG 3306
Db 3415 TGCATCTTTGCCAAGAGAACAGAGAGTGAATAATAGAAAAATATGCGATGTATATCG 3474
QY 3307 GCTGTGTCCAGAAATCTACTGCTTAAACCAGACAGAGGAGGAAAAACCATATGAATCTCA 3366
Db 3475 GCATGTGTGAGAAATCTACCGCCGCTTAACCCAGACAGAGGAGGAAAGCTGTGACACTTA 3534
QY 3367 GGAGAAAGCCCTAGACCCATTTTCAATGATTCAGACTTGGCATATGCAATTAACAGGA 3426
Db 3535 GGAGAAAGCCCTAGACCCATTTTCAATGATTCAGACTTGGCATATGCAATTAACAGGA 3594
QY 3427 AGCTGTGTATGTAAATGCAACGAGTGTGCTGGGAGAAATTTTGAAGCTGTACAGCTG 3486
Db 3595 AGCTGTGTATGTAAATGCAACGAGTGTGCTGGGAGAAATTTTGAAGCTGTGCAAGCTG 3654
QY 3487 AGCTGTGAGAGGCAATTCATGTTGACCTTTTACCTTTGGAAGTGGAGATATCTTTGC 3546
Db 3655 AGCTGTGAGAGGCAATTCATGTTGACCTTTTACCTTTGGAAGTGGAGATATCTTTGC 3714
QY 3547 CCTCTTTGCAAAATCTGTGCAATACTGTATGCCCATTTATTCCTTTGCAACTCAAAAG 3606
Db 3715 CCGCTGTGCAAGTCTGTCTGCAACACTGTCAATCCCATTCATCCTTTGCAACCGAGAG 3774
QY 3607 ATTAACAGTGAATGACAGATGCTCTTGCTCACTTTTGAACCTTGGCACGCTGATACAG 3666
Db 3775 ATCAACAGTGAATGAGAGGCTGCTTGCATCACTTTTGAACCTTGGCCGCTGATACAG 3834
QY 3667 ACTGTGTGGCGAATATGAGTTATATATAGATGATGTAAGGAGAAAC--CA 3723
Db 3835 ACTGTGTGGCGAATATGAGTTATATATAGATGATGTAAGGAGAAAC--CA 3894
QY 3724 ATTCATATTTCTTTATCAAGAGATGGAGATTTCTATTTGAGATTTCATTCATCTCG 3783
Db 3895 GTTCCTGTCTTTTATCAAGAGATGGGAGATTTCACTTTGAGTTTCATTCATCTCG 3954
QY 3784 AGTTTGGGCTGAGTCTTGATTAATTAATTAATAGCATCAAGAAAGGTTATCTC 3843
Db 3955 AGTTTGGGCTGAGTCTTGATTAATTAATTAATAGCATCAAGAAAGGTTATCTC 4014
QY 3844 TTGSCACAACATTTATATAGATTTGATTAAGTGGCACCTGCACTGTAAGAGGATTCCTG 3903
Db 4015 TTGSCACAACATTTATACAAATTTGAGTGAAGTGGCACCTGCACTGTAAGAGGATTCCTG 4074
QY 3904 GTCCCATGCTGACCTGAGACACCTGCGCTTTCATCTCAAGCAATTAAGAAATCTATG 3963
Db 4075 GTCCCATGATGACCTGAGACACCTGCGCTTTCATCTCAAGCAATTAAGAAATCTATG 4134
QY 3964 GGAGATGAAGGAAACCTCTGTTTGAAGCACTTCAAAATAGGACATATATGCTGAAA 4023
Db 4135 GGAGATGAAGGAAACCTCTATTTGGAGCACTTCAAAATAGGACATATATGCTGAAA 4194
QY 4024 GCATTAATGAGTTTGAGCTGACAGAGAGATTAACCTGTCTCAGGCTCTATATACAGAA 4083
Db 4195 GCGCTTAATGAGTTTGAGCTGACAGAGAGATTAACCTGTCTCAGGCTCTATATACAGAA 4254
QY 4084 CATCTGCTGCTGCTTCTATAGTGTCTTCTCAACATTAATAGAGATGAGATACACATGC 4143
Db 4255 CATCTGCTGCTGCTCTCTGATGATTTCTCTTACCTGCAATAGAAAAATACACAGGC 4314
QY 4144 CTTCCTGTATAGATCTGTTTCAATGTTTGGTGGGCTGCTGTTAGCATTCATCTTG 4203
Db 4315 CTTCCTGTGATGATCTCTTCAATGTTTGGTGGGCTGCTGTTAGCATTCATCTTG 4374
QY 4204 TATGAGATGACCTGTTGATGTGACGCTTCTCAATGTTCTTCCATTAACACCTT 4263
Db 4375 TATGAGATGACCTGTTGATGTGACGCTTCTCAATGTTCTTCCATTAATTAACACCTT 4434
QY 4264 TATCTCTTCATTTGATCAACATGAGCACAATGCTTAAGATACACTTAACGATGACA-- 4321
Db 4435 TATCTCTTCATTTGATCAACATGAGCACAATGCTTAAGATACACTTCATTAACGATGACA 4494

OY	689	TACAAAGAGCTTGACTGTAGCTCGAGAGGCCAGTTGCTATACCACTGCCATTGACA	748
Db	241	TACAAAGAGCTTGACTGTAGCTCGAGAGGCCAGTTGCTATACCACTGCCATTGACA	300
OY	749	AAGAGGCTCCGGGCTGTAAAGCGGAGGCTTATGCTGTGGCAGGAGCAAGAGAG	808
Db	301	AAGAGGCTCCGGGCTGTAAAGCGGAGGCTTATGCTGTGGCAGGAGCAAGAGAG	360
OY	809	ATATTAAGAGCATTCAGAAAATGCTCTCAACATCCACTTCATGTAGAACTATTACCT	868
Db	361	ATATTAAGAGCATTCAGAAAATGCTCTCAACATCCACTTCATGTAGAACTATTACCT	420
OY	869	CAGAGATTATGGCTCATCAGAAAATTTGGCTTGCGCTGGTCCGATGGAACAAATTA	928
Db	421	CAGAGATTATGGCTCATCAGAAAATTTGGCTTGCGCTGGTCCGATGGAACAAATTA	480
OY	929	TGAGCTATTCAAGTGACTTTAGGAGAGATCTTTGCCAAGCATGCGCTTAGAGAGAACTG	988
Db	481	TGAGCTATTCAAGTGACTTTAGGAGAGATCTTTGCCAAGCATGCGCTTAGAGAGAACTG	540
OY	989	ACTGGGAGATCCCTGTCTCATTAAGCAGGTTAATGCTTTGGGATGCAAGCTTTATAAG	1048
Db	541	ACTGGGAGATCCCTGTCTCATTAAGCAGGTTAATGCTTTGGGATGCAAGCTTTATAAG	600
OY	1049	GTGCCCGTAATCCTTCATGAATTTGATGATCCAGAGTTTATTAAGAGATGGAATACA	1108
Db	601	GTGCCCGTAATCCTTCATGAATTTGATGATCCAGAGTTTATTAAGAGATGGAATACA	660
OY	1109	AAAAACTCTTGTGATGGAATTTTGAAGTATTAATAACAATCAGAAATAATATACA	1168
Db	661	AAAAACTCTTGTGATGGAATTTTGAAGTATTAATAACAATCAGAAATAATATACA	720
OY	1169	GTGATGATCATGACAGAAATATCTCTATTACTGCACCTTCAGTTACAGATTTACTGTTG	1228
Db	721	GTGATGATCATGACAGAAATATCTCTATTACTGCACCTTCAGTTACAGATTTACTGTTG	780
OY	1229	CTACTCGGCTCGACATCTTTATTAAGAGCAAGATGTATCTGCTCATTTCTGAAATC	1288
Db	781	CTACTCGGCTCGACATCTTTATTAAGAGCAAGATGTATCTGCTCATTTCTGAAATC	840
OY	1289	TGCTAGAAAGTTTACTGTAGTACTTGAGCAGAGAAATAAATTAACCTCAGGGTATTA	1348
Db	841	TGCTAGAAAGTTTACTGTAGTACTTGAGCAGAGAAATAAATTAACCTCAGGGTATTA	900
OY	1349	GCCAGGACAAATTTGGAGAGTATATGCAATATGTGACCTAAAGTATATCTGTATCA	1408
Db	901	GCCAGGACAAATTTGGAGAGTATATGCAATATGTGACCTAAAGTATATCTGTATCA	960
OY	1409	GCAAAACCCACAATATGACAGAGAAAGATTAGAATGCACTTCTCTGAAGGTTTGATCTT	1468
Db	961	GCAAAACCCACAATATGACAGAGAAAGATTAGAATGCACTTCTCTGAAGGTTTGATCTT	1020
OY	1469	TTTTGAAGATTTTACCTGATGAGAGGAATGGAAGAAATCCGAAGACGTTGGGCAC	1528
Db	1021	TTTTGAAGATTTTACCTGATGAGAGGAATGGAAGAAATCCGAAGACGTTGGGCAC	1080
OY	1529	ACATTGAAGTGATCTGATTTGGGAGGCTGCCATTGCTATACAGATGCAATTTGAAGATA	1588
Db	1081	ACATTGAAGTGATCTGATTTGGGAGGCTGCCATTGCTATACAGATGCAATTTGAAGATA	1140
OY	1589	TTTTACTCATGTTCCAAAGAGTGTGTGCTGTGATGAAGAACTCTTACTTGCTTATA	1648
Db	1141	TTTTACTCATGTTCCAAAGAGTGTGTGCTGTGATGAAGAACTCTTACTTGCTTATA	1200
OY	1649	AAGAAATGCAAAAGCTGTGATGTAGAGTCCACTACCAAGTTTCATATCTAGTAGCAAGAC	1708
Db	1201	AAGAAATGCAAAAGCTGTGATGTAGAGTCCACTACCAAGTTTCATATCTAGTAGCAAGAC	1260
OY	1709	TAGTACAAATCTGTGGAGACATAGTTTGGAAACAAAGTCTTACAGAGTATCTAGAGATCTTG	1768
Db	1261	TAGTACAAATCTGTGGAGACATAGTTTGGAAACAAAGTCTTACAGAGTATCTAGAGATCTTG	1320
OY	1769	TAAACATACATCTGCCACTCTCTAGAGACCTTGTGCTGCTTCATGATAGTTTAAAGCAGGC	1828

Db	1321	TAAGATACATCTGCACCTCTCTAGAGACCCTTGCTGCTTCATGTACGTTAAAGACGCC	1380
OY	1829	TGGGTGCTGTTTCAAGACATGCATCAATTTTGTCTTTTGGAGACCTTTCCAAATAGAGTAC	1888
Db	1381	TGGGTGCTGTTTCAAGACATGCATCAATTTTGTCTTTTGGAGACCTTTCCAAATAGAGTAC	1440
OY	1889	TAGTGGAAATATCCCTTTACGTGTGGGTGGTGGTCCGACAGTGTGGCTAGATGTGGC	1948
Db	1441	TAGTGGAAATATCCCTTTACGTGTGGGTGGTGGTCCGACAGTGTGGCTAGATGTGGC	1500
OY	1949	GAAGAAATGAGCTGCTCTTATTTATAGCCAGGTGTTTATTTACCAAGATTTTAAAGTCAGAG	2008
Db	1501	GAAGAAATGAGCTGCTCTTATTTATAGCCAGGTGTTTATTTACCAAGATTTTAAAGTCAGAG	1560
OY	2009	AAGAAATGTATGATTAAGATATTCATCATGCTTCAGATTGTGCATCTTTAAATGATCCCA	2068
Db	1561	AAGAAATGTATGATTAAGATATTCATCATGCTTCAGATTGTGCATCTTTAAATGATCCCA	1620
OY	2069	ATAAGTCTGTTTACTGCTACTTCAGAGGTATGAACTTCCGAGAGCTTTTAAACAAGNCCA	2128
Db	1621	ATAAGTCTGTTTACTGCTACTTCAGAGGTATGAACTTCCGAGAGCTTTTAAACAAGNCCA	1680
OY	2129	TATCTACAAAGACAGAGATTGATTTAAACAATTTAATATACACTATATAGAAATGCTTC	2188
Db	1681	TATCTACAAAGACAGAGATTGATTTAAACAATTTAATATACACTATATAGAAATGCTTC	1740
OY	2189	AGGTCCTCATCTATATTGTGGGTGAGCGTTATGTACTCGAGTGGGAATGTGACCAAG	2248
Db	1741	AGGTCCTCATCTATATTGTGGGTGAGCGTTATGTACTCGAGTGGGAATGTGACCAAG	1800
OY	2249	AAGAGGTACAAATGAGAGAAATCATCTACCTTGCTTTGCAATTGAACCCATGCCACACAGT	2308
Db	1801	AAGAGGTACAAATGAGAGAAATCATCTACCTTGCTTTGCAATTGAACCCATGCCACACAGT	1860
OY	2309	CCATTTGCCAAAATTTTACCTGAGATGAAATTAATGAACACTGGCTTAGAGATGTCAATA	2368
Db	1861	CCATTTGCCAAAATTTTACCTGAGATGAAATTAATGAACACTGGCTTAGAGATGTCAATA	1920
OY	2369	ACAAAGTGGCCACATTTAAGAAACACAGGTGTATCAGGCCATGGAATTTATGACTAAAG	2428
Db	1921	ACAAAGTGGCCACATTTAAGAAACACAGGTGTATCAGGCCATGGAATTTATGACTAAAG	1980
OY	2429	ATGATATCACTGAAGACTTCAATATGTATCTTTTATCTACTCCAAAACCCAGATAGCA	2488
Db	1981	ATGATATCACTGAAGACTTCAATATGTATCTTTTATCTACTCCAAAACCCAGATAGCA	2040
OY	2489	AGCGTGAACATATGACAGAGAAAGGAGAAACAAGAAACAAAGATGAACATTTGCCGC	2548
Db	2041	AGCGTGAACATATGACAGAGAAAGGAGAAACAAGAAACAAAGATGAACATTTGCCGC	2100
OY	2549	CACCAACACCTCTCGAAATTTGCCCCGTGTTTACACAAAGTCAATTAACCTTCAACTGTG	2608
Db	2101	CACCAACACCTCTCGAAATTTGCCCCGTGTTTACACAAAGTCAATTAACCTTCAACTGTG	2160
OY	2609	ATATCATGATGTACATTTCTCAGAACCCGATTTTGAAGCGGCAATFAGACACAGATTTCACT	2668
Db	2161	ATATCATGATGTACATTTCTCAGAACCCGATTTTGAAGCGGCAATFAGACACAGATTTCACT	2220
OY	2669	TGTGGACCGAAGGATGCTCCAAATGTGCTTTTCAATTTCTGGCAATTTGGGTTTACTAGAG	2728
Db	2221	TGTGGACCGAAGGATGCTCCAAATGTGCTTTTCAATTTCTGGCAATTTGGGTTTACTAGAG	2280
OY	2729	AGAAGCAACACCTTCAAAAACCTCCTGGAAGAAGATTAACATTTTGAATCTTTATCATTAAG	2788
Db	2281	AGAAGCAACACCTTCAAAAACCTCCTGGAAGAAGATTAACATTTTGAATCTTTATCATTAAG	2340
OY	2789	CTTCAGATTTGGAGATTCACGCATATGATATACAAATGCTTTTGGAAAACTCAAAGAA	2848
Db	2341	CTTCAGATTTGGAGATTCACGCATATGATATACAAATGCTTTTGGAAAACTCAAAGAA	2400
OY	2849	TTCCCCAGTTAGAAAGCCAGAAGACATGATTAACGTGATTACTTACATGTTTGACACAG	2908

Db 2401 TTCCCGAGTTAGAGCCAGAGGACATGATACGATGATACCTGATACCTTACAGATGTTTGACACAG 2460

QY 2909 TGAAGCGATTAGAGAAAATCTTGTAAATTTAGCAACCAACCATCAGGATCGGAATCTA 2968

Db 2461 TGAAGCGTTTGAAGAAAATCTTGTAAATTTAGCAACCAACCATCAGGATCGGAATCTA 2520

QY 2969 TTAAGAAATGATGAGATTACTCATGATAAAG 2998

Db 2521 TTAAGAAATGATGAGATTACTCATGATAAAG 2550

RESULT 5

ABQ75898

ID ABQ75898 standard; cDNA; 1635 BP.

XX

AC ABQ75898;

XX

DT 17-OCT-2002 (first entry)

XX

DE Human ubiquitin relative protein 46.64 cDNA.

XX

KW Human; ubiquitin relative protein 46.64; tumour; inflammation;

KW immunological disease; haemopathy; human immunodeficiency virus; HIV;

KW gene; ss.

XX

OS Homo sapiens.

XX

FH Key

FT 55..1329

FT CDS

FT /*tag= a

FT /product= "ubiquitin relative protein 46.64"

XX

PN CN1339485-A.

XX

PD 13-MAR-2002.

XX

PF 23-AUG-2000; 2000CN-0119708.

XX

PR 23-AUG-2000; 2000CN-0119708.

XX

PA (BODE-) BODE GENE DEV CO LTD SHANGHAI.

XX

PI Mao Y, Xie Y;

XX

WPI; 2002-464069/50.

DR

DR P-PSDB; ABQ75898.

XX

PT New polypeptide-human ubiquitin relative protein 46.64 for

PT treating malignant tumours, inflammations, immunological diseases,

PT haemopathy and human immunodeficiency virus infection -

XX

PS Claim 6; Page 24-25 (disclosure); 33pp; Chinese.

XX

CC The present invention discloses a new kind of polypeptide, human

CC ubiquitin relative protein 46.64, polynucleotides for encoding this

CC polypeptide, and a DNA recombination process to produce the polypeptide.

CC The present invention also discloses the method of applying the

CC polypeptide in treating various diseases, such as malignant tumours,

CC inflammations, immunological diseases, haemopathy and human

CC immunodeficiency virus (HIV) infection. The current sequence represents

CC the human ubiquitin relative protein 46.64 cDNA.

XX

SQ Sequence 1635 BP; 529 A; 290 C; 344 G; 472 T; 0 other;

Query Match 25.2%; Score 1311; DB 24; Length 1635;

Best Local Similarity 99.2%; Pred. No. 0;

Matches 1317; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1387 GACCTAAATATATCTGTATGATGACCAACCCCAATATGACAGAAAGATTAAAGATGCAG 1446

Db 1 GACCTAAATATATCTGTATGATGACCAACCCCAATATGACAGAAAGATTAAAGATGCAG 60

QY 1447 TTCCTTGAAGGTTTTCGATCTTTTGAAGATTCTTACCTGTATGACGGGAATGGAAGAA 1506

Db 61 TTCCTTGAAGGTTTTCGATCTTTTGAAGATTCTTACCTGTATGACGGGAATGGAAGAA 120

QY 1507 ATCCGAAGACAGGTTGGGCAACACATTTGAAGTGGATCCTGATTTGGAGGCTGCCATTCCT 1566

Db 121 ATCCGAAGACAGGTTGGGCAACACATTTGAAGTGGATCCTGATTTGGAGGCTGCCATTCCT 180

QY 1567 ATACAGATGCAATTGAAGAATATTTTACTCATGTTCCCAAGAGTGGTGTCTGTGATGAA 1626

Db 181 ATACAGATGCAATTGAAGAATATTTTACTCATGTTCCCAAGAGTGGTGTCTGTGATGAA 240

QY 1627 GAACCTTTACTTTGGGCTTATAAAGAATGTCACAAAGCTGTGATGAGTGCAGTACCAGT 1686

Db 241 GAACCTTTACTTTGGGCTTATAAAGAATGTCACAAAGCTGTGATGAGTGCAGTACCAGT 300

QY 1687 TTCATATCTAGTAGCAAGACAGTAGTACAATCGTGTGGACATAGTTTGGAAACAAGTCC 1746

Db 301 TTCATATCTAGTAGCAAGACAGTAGTACAATCGTGTGGACATAGTTTGGAAACAAGTCC 360

QY 1747 TACAGAGTATCTGAGGATCTTTGTAAGCATACATCGCCACTCTCTPAGGACCTTCTGCTGGT 1806

Db 361 TACAGAGTATCTGAGGATCTTTGTAAGCATACATCGCCACTCTCTAGGACCTTCTGCTGGT 420

QY 1807 CTTTCATGTACGTTTAAAGCAGGCTGGGTGCTGTTTCAAGACTGCATGAATTTGTGCTTTT 1866

Db 421 CTTTCATGTACGTTTAAAGCAGGCTGGGTGCTGTTTCAAGACTGCATGAATTTGTGCTTTT 480

QY 1867 GAGGACTTTTCAAGTAGAGGTACTAGTGAATATCTTTACGTTGTCTGGTGTGCTGGTTCGC 1926

Db 481 GAGGACTTTTCAAGTAGAGGTACTAGTGAATATCTTTACGTTGTCTGGTGTGCTGGTTCGC 540

QY 1927 CAGGTTGTGCTGAGATGTGCGGAAGAAATGACTGTCTCTTATTAGGCAGGTGTTTAT 1986

Db 541 CAGGTTGTGCTGAGATGTGCGGAAGAAATGACTGTCTCTTATTAGGCAGGTGTTTAT 600

QY 1987 TACCAAGATGTTAAGTGCAGAGAGAAATGTATGATAAAGATATCATGCTTCAGATT 2046

Db 601 TACCAAGATGTTAAGTGCAGAGAGAAATGTATGATAAAGATATCATGCTTCAGATT 660

QY 2047 GGTGCATCTTTAATGGATCCCAATAGTCTTGTGTACTTGGTGTGCTGAGGATGATCACTT 2106

Db 661 GGTGCATCTTTAATGGATCCCAATAGTCTTGTGTACTTGGTGTGCTGAGGATGATCACTT 720

QY 2107/GCCGAGGCTTTTAAACAAGACCATATCTACAAAGACCAAGGATTTCATTAAACAATATAT 2166

Db 721 GCCGAGGCTTTTAAACAAGACCATATCTACAAAGACCAAGGATTTCATTAAACAATATAT 780

QY 2167 ACCTAATAGAGAAATGCTTTCAGGTCTCTCATCTATATTGTTGGTGTGAGCGTTATGACCT 2226

Db 781 ACCTAATAGAGAAATGCTTTCAGGTCTCTCATCTATATTGTTGGTGTGAGCGTTATGACCT 840

QY 2227 GGAGTGGGAATGTGACCAAGAGAGGTACATGAGAGAAATCATTCACCTTGTCTTGC 2286

Db 841 GGAGTGGGAATGTGACCAAGAGAGGTACATGAGAGAAATCATTCACCTTGTCTTGC 900

QY 2287 ATTCAACCCATGCCACACAGTGCCATTGCCAAAATTTACCTGAGAATGAAAATATGAA 2346

Db 901 ATTCAACCCATGCCACACAGTGCCATTGCCAAAATTTACCTGAGAATGAAAATATGAA 960

QY 2347 ACTGGCTTAGAGAATGTCTATAAACAAGTGCCCATTTTAAAGAACCAAGGTTATCAGGC 2406

Db 961 ACTGGCTTAGAGAATGTCTATAAACAAGTGCCCATTTTAAAGAACCAAGGTTATCAGGC 1020

QY 2407 CATGGAGTTTATGAACCTAAAAGATGAATCTACTGAAAGACTTCAATATGACTTTTATCAT 2466

Db 1021 CATGGAGTTTATGAACCTAAAAGATGAATCTACTGAAAGACTTCAATATGACTTTTATCAT 1080

QY 2467 TACTCCAAAACCCAGCATAGCAAGGCTCAACATATGACAGAGAAAGAGAGAAACAGAA 2526

Db 1081 TACTCCAAAACCCAGCATAGCAAGGCTCAACATATGACAGAGAAAGAGAGAAACAGAA 1140

QY 2527 AACAAAGATGAAGCATTTGCCGCCACCACTCTCTGAAATCTTGCCTTCTGCTTTCAGCAAA 2586

Db	1141	AACAAGATGAAGCATTTGCCGCGCACACACACACTCCTCTGTAATTCCTGCTTCACGAAA	1200
Qy	2587	GTGATTAACCTTCTCAACTGTGATATCATATGATGACATTTCCAGAGACCGATTTGACGG	2646
Db	1201	GTGATTAACCTTCTCAACTGTGATATCATATGATGACATTTCCAGAGACCGATTTGACGG	1260
Qy	2647	GCAATTAACACAGATTTCTAATCTGTGGACCGAGAGATGCTCCAAATGGCTTTTCATATT	2706
Db	1261	GCAATTAACACAGATTTCTAATCTGTGGACCGAGAGATGCTCCAAATGGATTTGGAGAGTT	1320
Qy	2707	CTGGCAT 2713	
Db	1321	CAGCCAT 1327	
RESULT 6			
AAK51709			
ID	AAK51709	standard; cDNA; 6840 BP.	
AC	AAK51709;		
XX			
DT	06-NOV-2001	(first entry)	
XX			
DE	Human polynucleotide seq	ID NO 254.	
XX			
KW	Human; cytokine; cell proliferation; cell differentiation; gene therapy;		
KW	vaccine; peptide therapy; stem cell growth factor; haematopoiesis;		
KW	tissue growth factor; immunomodulatory; cancer; leukaemia;		
KW	nervous system disorder; arthritis; inflammation; ss.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200157190-A2.		
XX			
PD	09-AUG-2001.		
XX			
PF	05-FEB-2001; 2001WO-US04098.		
XX			
PR	03-FEB-2000; 2000US-0496914.		
PR	27-APR-2000; 2000US-0560875.		
PR	20-JUN-2000; 2000US-0598075.		
PR	19-JUL-2000; 2000US-0620325.		
PR	01-SEP-2000; 2000US-0654936.		
PR	15-SEP-2000; 2000US-0663561.		
PR	20-OCT-2000; 2000US-0693325.		
PR	30-NOV-2000; 2000US-0728422.		
XX			
PA	(HSE-) HSEQ INC.		
XX			
PI	Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;		
PI	Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;		
PI	Xue AJ, Yang Y, Wejhrman T, Goodrich R;		
XX			
DR	WPI: 2001-476283/51.		
DR	P-PSDB; AAM78576.		
PT			
PT	Nucleic acids encoding polypeptides with cytokine-like activities,		
XX	useful in diagnosis and gene therapy -		
PS	Claim 1; Page 1159-1166; 6221pp; English.		
XX			
CC	The invention relates to polynucleotides (AAK51456-AAK53435) and the		
CC	encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to		
CC	cytokine, cell proliferation or cell differentiation or which may induce		
CC	production of other cytokines in other cell populations. The		
CC	polynucleotides and polypeptides are useful in gene therapy, vaccines or		
CC	peptide therapy. The polypeptides have various cytokine-like activities,		
CC	e.g. stem cell growth factor activity, haematopoiesis regulating		
CC	activity, tissue growth factor activity, immunomodulatory activity and		
CC	activin/inhibin activity and may be useful in the diagnosis and/or		
CC	treatment of cancer, leukemia, nervous system disorders, arthritis and		
CC	inflammation.		
CC	Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666		

CC	xx	Sequence	6840 BP; 2044 A; 1311 C; 1450 G; 2035 T; 0 other;	
CC	xx	Query Match	21.1%; Score 1097.8; DB 22; Length 6840;	
CC	xx	Best Local Similarity	55.9%; Pred. No. 1.6e-282;	
CC	xx	Matches 2408; Conservative	0; Mismatches 1807; Indels 96; Gaps 13;	
CC	xx	we're missing at the time of publication.		
CC	xx	(AAM80020) are omitted as the relevant pages from the sequence listing		
CC	xx	were missing at the time of publication.		
QY	942	TGACTTATAGCAGATCTTTTGGCCAGACATGCGTTAGAGAAAGACCTGACTCGGAAATCC	1001	
DB	217	TGGCCTTCGCCGGATTTTATGTGCAAGTTGGTTTACAGAAAGGCCAGATGGTGAACATC	276	
QY	1002	CTGTCTCATATAGCAGGTTAATGCTTTGGGATGCGAAAGCTTTATAAGGTGGCCCTAAGAT	1061	
DB	277	TTCCTCAGTGCACGACGTAGTACTTATGATTCGAAATTAAGGAAAGGTGCTAGAGATGT	336	
QY	1062	CCCTTCATGAATTTGATCTTCAGACAGCTTTTATTTATGAGANTGGATACAAAACCTCTTGC	1121	
DB	337	ATATCATCAGTGTTCATGACAGCTGCTTATGATTTGAAATACAGAAACCTATTTGC	396	
QY	1122	TATGGAATTTGTGAGATTTATAAACAATGCGAGAAAGAAATATATCAGTATGATCATGA	1181	
DB	397	TGTTGATTTGCCAAAAAACAATATGAGCGCTTTCAGAGATGATATATGACAGATGACACGA	456	
QY	1182	CAGAAATATCTCTATAACTGCATTTAGTTACAGATGTTTACTGTCTCTACTGCGTCG	1241	
DB	457	CAGAGAGTTTTCACATGCTCCAGACCTCTGCGTTCAATATTCACGGTTCCTTCACCTGCTCG	516	
QY	1242	ACATCTTATATGAGAGAGAGAAATGTTATCTCTGATTAACAGTAACTGCTAGAAAGTTT	1301	
DB	517	AATGCTCATCAGACAGAAAGAAACCTTGATGAGCATTTATATTAAGCTTTTATGATTCATTT	576	
QY	1302	A-----CTGAGTACTTGGACAGAGAAACAATAAATTCACCTTCAGGGGTTATAGCCAGGA	1355	
DB	577	GAGACATCGAGATCCCGAGGCGACATTTTCACTTTGAACGATACACGCTGTTTACAGCCTT	636	
QY	1356	CAAAATGGGAAGATATATGCAATATATGACCTTAAGATATCCTGATCAGCAACAC	1415	
DB	637	CAAAATTTAAGGAGATACAGACCCTTATTTTATAGATCTCAAGATGTGTTTAATTAACCAAC	696	
QY	1416	CACAAATTTGACAGAAAGATTAACAAATGCACTTCCTTGAAGCTTTGCATTTTGTGAA	1475	
DB	697	AACGTAAATGTCACATGACGTAGGACAGAGAGTTCCTTAGAAGGGTTTATGCTTTTGA	756	
QY	1476	GATTCCTTACCTGTATGACAGGGAAATGGAAGAAATCCGAAAGACAGGTGGGCAACATGGA	1535	
DB	757	ATTACTTAAATATGTATGACAGGGAAATGATCCAAATTAACGTCAGTAGAGCAACAATGGA	816	
QY	1536	AGTGAATTCGTATTTGGAGGCTGCCATTTGCTATACAGATGCAATTTGAAGATATATTTACT	1595	
DB	817	AATGGAACCAAGAGTGGGGAAGCAGCCTTCACCTCAAAATGAATTAACATCATGCAATTC	876	
QY	1596	CATTTCCAAAGAGTGTGCTCTTGATGGAAGAACTCTTACTTGGGCTTATAAGATG	1655	
DB	877	AATCATGAGAGACTGGTGTGCTTAGATGAAAAAGTGTATTCGAAAGCTTTACAGAAATG	936	
QY	1656	TCACAAAGCTGTATGAGTGCACGTATACCAGTTTATATCTAGTAGCAAGACAGT---AGT	1712	
DB	937	TCTGGCTGTACTGATGCAGTGTCAATGATGGTTATACAGATGATGTAACGCCAATACACACT	996	
QY	1713	ACAATTCGTGTGACATAGTTTGGAAACAAGCTCCTACAGAGTATCTGAGATCTTTGTAAG	1772	
DB	997	AAGCATTTGTGACATTCAGTGGAAACATACAGATACGTGTTTCCCAAGAAAAAGTTAG	1056	
QY	1773	CATCAATTCGCACATCTCTAGACCCCTTGCGATTCATCATAGCTTAAACACAGGCTGGG	1832	
DB	1057	CATTACCTCCCAAGTTTCTCGCTTACCTTGCAAGTTTACATATATTTATTAACCAAAAGTGA	1116	
QY	1833	TGCGTTTTCAGACATGCATGAATTTGTTGCTCTTTGAGAGACTTTCAAAGTAGAGTACTAGT	1892	
DB	1117	AGTGGCATATATAATTTCCAGAGCTCTACTCTTAAGTAGAAGTACCCACCCAGTGTGAT	1176	

1893 GGAATATCCTTTACGTTGCTGCTGTTGGTTGGCTGCCAGGTTGTTGCTGAGATGTGGCGAAG 1952
1177 AGAACCCCTCTTTAGATGCTCTGTTGTTGGTCCCAAGTACATGCCGGAATGTGGAGAG 1236
1953 AAATGGACTGCTCTTTATTACCCAGGTGTTTATTACCAAGATGTTAAAGTGCAGAGAAGA 2012
1237 AAATGGGTTCTCTAGTAACCCAGATTTATTAATACCATTAATGTGAATGCAGACGTGA 1296
2013 AATGTATGATAAGATATCATATGTTTCAGATTTGGTGCAATCTTTAATGATGCCAATAA 2072
1297 GATGTTTGAAGAGATGTAGTAATGCTTCAGACAGGTGTCTCCATGATGGATCCAAATCA 1356
2073 GTTCTGTGTTACTTACAGAGGTATGAATGCTGCGAGGCTTTTAAACAGACCATATC 2132
1357 TTTTCTGTGATCATGCTCAGCGCTTTGAACCTTTATCAATTTTTCAGATTTTTCAGACTA 1416
2133 TACAAAAG-----ACCAGGATTTGATTAACAAATATATACACT 2171
1417 TGGAAAAGATTTAGTTCTGAGATTACCCATAGGATGTTGTTTCAGCAGAACATATCTCT 1476
2172 AATAGAAGAAATGCTTCAGTCCCTCAATCTATATTGTTGGGTGAGCGTTTATGTACCTGGAGT 2231
1477 AATAGAAGAAATGCTATACCTCATTAATAATGCTTGTGTTGGAGAGAGATTTAGTCCCTGGAGT 1536
2232 GGGAAATGTACCAAGAGAGGTCAATGAGAGAAATCATTCACCTGCTTTCATGTA 2291
1537 TGGACAGGTAAATGCTACAGATGAATCAAGCAGAGATTAATCCATCAGTTGAGTATCAA 1596
2292 ACCATGCCACACAGTCCCATGCGAAAATTTACCTGAGAATGAAAATATGAACATGG 2351
1597 GCCTATGGCTCATAGTAATTTGGTTAAAGTCTTTACCTGGAATGAGAACRAGAGACTGG 1656
2352 CTTAGAGAAATGCTATRAACAAAGTGGCCACATTTAAGAAACCGAGTGTATCAGCCCATGG 2411
1657 CATGGAGAGTAAATCGAAGCAGTGGCCATTTCAAGAACTTGGATTAACAGGACGAGG 1716
2412 AGTTTATGAACTAAAGATGAATCAGTGAAGACTTCAATATGATGACTTTTATCATTAATC 2471
1717 CATGTATGAATGAACAGCAATGTGCCAAGAGTTCAACTTGTATTTCTATCACTTTTC 1776
2472 CAAAACCCACATAGCAAGCTGAACATATGCAAGAAAGAGGAGAAACAAAGAAACAA 2531
1777 AAGGCGAAGACAGTCCAAAGCAGAAAGCGCACGGAATTTGAAAGACAAATATAGAGA 1836
2532 AGATGAGACATGTGCGCCACACACCTCCTGATTTCTGCCCCCTTTTCAGCAAGTAT 2591
1837 AGATACAGACTCCACCTCCGGTGTGCTCCATTCCTGCTGCTGTTTGAAGCCTGGT 1896
2592 TAACCTTCTCACTGTGATATCATATGATGATGATGATGATGATGATGATGATGATGATGAT 2651
1897 TAACATTTTGCAGTCAGATGTCATGTTGTGTCATCATGGGAACAAATTCGCAATGGCTGT 1956
2652 AGACACAGATTTCACTTTGTCGCGAAGGATGCTCCAAATGCTTTTTCATATCTCTGGC 2711
1957 GGAACATAATTTGGATATGCTTGGTCAGAGTCCATGCTGCAAGGGGTGTACATTTAATTTGG 2016
2712 ATTGGGTTTACTAGAGAGAGAACAGCTTCAAAAAGCTCCTGAAG---AAGAAGTAAAC 2768
2017 CATGGCACTACAGAGAAACAAACATTTAGAGAAATGTACGGAAGCAATGTAGTAAC 2076
2769 ATTTGACTTTTATCATAGGCTTCAAGATTTGGGAAGTTCAGCCCATGAAT-----AT 2819
2077 ATTTACCTTCTACATCAGAGATATCAAAACCTGGTGAAGCGCCAAAATTTCTCTAGCAT 2136
2820 ACAATGCTTTTGGAAAACATCAAGAAATTTCCCGAGTTAGAAGGCCAGAGGACATGAT 2879
2137 ACTAGCTATGCTGGAAACACTACAAAATGCTCCCTACCTAGAAGTCCACAAAGACATGAT 2196
2880 AACTGTGATCTTCCAGATGTTTGCACAGTGAAGCGATTAAGAGAAAATCTCTGTTTAAT 2939
2197 TCGGTGGATATTTGAAGACTTTTAAATGCTGTTTAAAGATGAGGAGA-----GTTACC 2250
2940 TGTAGCAACCATCATCAGGATCGGAATCTATTAGAATGATGAGATTACTCATGATAAGA 2999

2251 TACCAGTCCCTGGCAGAGACAGAGGAGCAACCAATATGAGAGAGAGTTCAAGGGACAAAGA 2310
3000 AAAAGCAGAGCAAAAGAAAGCTGTAAGCTGCTAGGCTACATCGCCAGGAAGATCATGCG 3059
2311 CAAAGCTGAGAGGAAGAAAGCAGAGATTGCCAGACTGCGCAGAGAAAGATCATGCG 2370
3060 TCAGATGCTGCTGTACAGAAAAACTTCAATGCAAACTCATAACTCATATGACATAC 3119
2371 TCAGATGCTGAAATGCGCGCATTTTATTGATGAAAACAAAGAACTCTTTTCAGCAGAC 2430
3120 ATCAGAAATGCTGGGAAGAAGATTCCTATATGAGGAAGAGAGACCCCGACGACTCAG 3179
2431 ATTAGAACTGGAGCTCAACTCTGCTGTTCTTGATCAT-----AGCCCTGTGGCTTC 2484
3180 TGACTACTCTAGAATTTGTTGGGTCTTAAAGGGGTCCATCTGTACTTGAAGAAAGGAGT 3239
2485 AGATATGACACTTACAGCACCTGGGCGCCGCAACAACTCAGGTTCTCTGAACAAAGACAAT 2544
3240 GCTGACGTGATCCTTTGCCAAGAAAGACAGAGGTGAAATAGAAAATAATATGCCATGCT 3299
2545 CGTTACATGATATTTGTTCAAGAGGAGCAAGTTTAAAGTGGAAAGCAGGCGCATGCT 2604
3300 ATTATCGGCTGTGTCAGAAATCTACTGCCCTTAACCCAGACACAGGGGAAAACCCATAGA 3359
2605 CTTGGCAGCATTTGTTTCAGAGATCAACTGTATTATCAAAAACACAGAAATTTATTCTCA 2664
3360 ACTCTCAGGAGAACCCCTAGACCCACTTTTTCATGATCCAGACTTGGCATATGGAACCTTA 3419
2665 AGATCCAGAAAA---TATGATCCATTTATCATGCCCTGATCTGCTTGTGGACACACA 2721
3420 TACAGGAAGCTGTGCTCATGTAATGACGAGTGTGCTGGCAGAGATATTTTGAAGCTGT 3479
2722 CACTAGTAGCTGTGGGCACATTTATGCATGCCCATTTGTTGGCAAGGATTTTGTATCCCT 2781
3480 ACAGCTGAGCTCTCAGCAGCGCATTCATGTTGACCTTTT-----TGACTTGA 3527
2782 TCAAGCTAAAGAACACGGAAGGCAACAGAGATTACGCTTTACATACAGAGTATGATGTAGA 2841
3528 AAGTGGAGAATATCTTTGCCCTCTTGTGCAATCTCTGTGCAATCTGTGATCCCCATTA 3587
2842 AAACGAGAAATCTTTTGGCCCTTTGTCCTGTAATGCTGTAATGCTGTTATTC 2895
3588 TCCTTTGCAACTCAAAAGATTAACAGTGAGAAATGCGAGATGCTGCTGCTCACTTTTGAC 3647
2896 TCTGCTGCTCTCCCAAGAAATATTTTAAACACAGGTTTAAATTTTTCAGACCAACAAA 2955
3648 CTTGCGCAGGTGGATACAGACTGTTCTGCCAGAAATATCAGGTTATAATATAAGACATGC 3707
2956 TCTGACTCAGTGGATTAGAACAAATATCTCAGCAAAATAAAGCAATTACAGTTTCTTAGGAA 3015
3708 TAAAGGAGAAAACCCCAATTCCTTATTTTCTTAAATCAAGGAATGGGAGATTTCTACTTTGA 3767
3016 AGAAGAAAGTACTCTTAATATGCTCTACAAGAAATTCAGAAATGTGGATGAATTACA 3075
3768 GTTCCATTCCTCCTGAGTTTGGGCTTGAGTCTTCGATTAATAATATTCAAATAGCATCAA 3827
3076 GCTCCCTGAAGGGTTCAGGCTGATTTTCGCTTAAGATCCCTTATTTCTGAGAGCATATA 3135
3828 GGAATGTTATTTCTTTTGGCAACAATTTATAGAAATTTGATTTGAAAGTGGCAGCTGA 3887
3136 AGAAATGCTTAACGACATTTTGGAACTGCTACCTACAGGTGGGACTAAAGGTTTCAATCCCA 3195
3888 TGAAGGATCCTCAGTCCCATCTGACCTGAGGACCTCGGCTTTTCACTATCCAGGC 3947
3196 TGAAGAGGATCCTCCTGTTTCCCATTAATGTTGGGGTAGCTGCGCGCTACACCATCCAAAG 3255
3948 AATTGAAATCTATTTTGGGAGATGAAGGAAAACCTCTGTTTGGGAGCACTTCAAATAGGCA 4007
3256 CATAGAAGAAATTTTGGAGTGAAGATAAACCATTTGTTGGTCTTTACCTTTGCGAGACT 3315
4008 GCATTAATGCTGAAGAGATTAATGAGTTTGCAGTTTGCAGAGGATTTACCTGTCCTCA 4067

Db 2371 CTCAGATGCTGAAATGACAGCGGCAATTTTATGATGAAACAAAGAACTCTTTCAGAGA 2430
 QY 3119 CATGAGAAATGCGTGGGAAAGAGATTTCATTATGAGAGAGAGAGCCAGCCAGCTCA 3178
 Db 2431 CATTAAGACTGATGCTCAACCTGCTGTTCTTGTATCAT-----AGCCGTGTGCTT 2484
 QY 3179 GTGACACTGCTAGAAATGCTTTGGTCTTAAAGGGGTCATCTGTTACTGAAAGAGAG 3238
 Db 2485 CAGATATGACACTTACAGCAGTGGGCCCGCACAACTCAGTTCTCGAACAAGACAAAT 2544
 QY 3239 TCGTGCAGTGCATCCTTTGCCAGAGAAACAGAGAGGAGTGAATAATATATGATCCATG 3298
 Db 2545 TCGTTCATGATGATATGTTGTGCAAGAGAGACAGAAATGTAAGAGAGAGGCAATGG 2604
 QY 3299 TATTATCGGCTGTGTCCAGAAATCTACTGCTTAAACCCAGCAGAGGGGAAACCCATAG 3358
 Db 2605 TCTTGGCAGCATTTGTTACAGAGATCAAGCTATATCAAAAAAACAAGAAATTTATTC 2664
 QY 3359 AACTTCAGAGAGAGCCCTAGACCCACTTTTCATGATCCAGACTTGGCATATGAACTT 3418
 Db 2665 AAGATCCAGAAAA---TAGATCCATTTATTCATGACACCTGATCTGTCTTGGAACAC 2721
 QY 3419 ATACAGAGAGCTGTGTGATGTAATGACAGAGTGTGCTGCGAAGATATTTGAACCTG 3478
 Db 2722 ACATAGTAGCTGTGGCACATTTGATGATGCCATTTGTGGCAAAAGTATTTGATTCG 2781
 QY 3479 TACAGCTGACCTTCAGCAGCGCATTCATGTCCTTT-----TGACTTGG 3526
 Db 2782 TTCAACCTTAAAGAACAGCGAAGGCAAGAGATTTAGCCTTACATAGAGCATATGATAG 2841
 QY 3527 AAAGTGGAGATATCTTTGCCCTTTGCAATCTCTGCAATCTGATCTGATCCCATTA 3586
 Db 2842 AAAACGAGAAATCTTTGCCCTTTGTAATGCTGATGATGATGATGATGATGATGATG 2896
 QY 3587 TTCTTTGCAACCTCAAAAGATTAACAGTAGATGATGATGATGATGATGATGATGATG 3646
 Db 2897 -TCTGCTGCTCTCCAGAAATATTTTAAACACAGTTAAATTTTTCAGACCAACCAA 2955
 QY 3647 CCTGACAGGCTGATACAGACTGTTCTGCCAGAAATATCAGGTTATTAATATAGACATG 3706
 Db 2956 ATCTGACTAGTAGGATTAAGAAATATCTAGCAAAATTAAGCATTTACATTTCTTGA 3015
 QY 3707 CTAAAGGAGAAAAACCAATTCCTATTTCTTAAACAAGAAATGGAGATTTCACTTTGG 3766
 Db 3016 AAGAAAGAACTACTCTTAATATGCTCTCAAAAGAAATCAAGAAATGTGATATATAC 3075
 QY 3767 AGTTCAATTCATCTGAGATTTGGGCTTGAATCTGATTAATATTAATATCAATATGATCA 3826
 Db 3076 AGCTCCCTAAGGGTTACAGGCTGATTTGCTCTTAAGATCCCTTATTTGAGAGCATAA 3135
 QY 3827 AGGAATGCTTATTTCTTTGCCACAACAATTTATGAATTTGATTAAGATGATGATGATG 3886
 Db 3136 AAGAAATGCTTAAGCAATTTTGAATGCTTACCTAAGAGTGGGCTTAAAGGCTTATCCCA 3195
 QY 3887 ATGAAAGGATCTCTGAGTCCCATGCTGACCTGAGACACCTGCTTCACTATCCAGG 3946
 Db 3196 ATGAAAGGATCTCTGAGTCCCATGCTGAGTGGGCTGAGTGGGCTGAGTGGGCTGAGT 3255
 QY 3947 CAATTAATATCTATTTGGGAGATGAAGAAACCTCTGTTTGGAGCACTTCAAAATAGGC 4006
 Db 3256 GCATTAAGAAATTTGAGTGAAGATTAAGCAATGTTTGGCTTTACCTTGAAGTGGAG 3315
 QY 4007 AGCATATGCTGCTGAAGCAATTAATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 4066
 Db 3316 TGGATGCTGCTTGAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 3375
 QY 4067 AGCTGCTGATGAGAAACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4126
 Db 3376 TTTCATGCTGAGAGAGCAATTTTGTAAACCTTTTGTGATGATGATGATGATGATGATG 3435
 QY 4127 CAGAGATACACCATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4186
 Db 3436 ATGAGAACTTTCATGATATTTAGATATGATGATGATGATGATGATGATGATGATGATG 3495

QY 4187 TAGCATTCACATCCCTGATTTGGGATGACCCCTGTGATCTGCAACCTTCTTCAATTAATT 4246
 Db 3496 TTGCATTTCTCGGCTTGGCAGTG-----TCAGATATTTTCAAGGATATCAGCC 3540
 QY 4247 CTTCCTATTAACACCTTATCTCTTCAATTTGATGACCTTGGCAGACATGCTGATACATG 4306
 Db 3541 TTGGCAGTGGAGACCTTACATTTTCCATCTGTTACTATGTCACACATCATACAGATCT 3600
 QY 4307 TACTTACAGTAGACAGAGCCCTACC---CCCTGCTCAGGTTCAAGAGACAGTGAAGAG 4363
 Db 3601 TACTTACCTCATATGATCAAGAGATGGATGATGATCAAGAAATCCCTTGTGAAGAG 3660
 QY 4364 CTCAATCCGCAATCTTTCTTTGCAAGAAATTTCAATATACAGTGGCTCCATTTGGGT 4423
 Db 3661 AATCAGCAGTCTTGTCTTTGATTAATAACACTTACACAGATATGCGGAGAGTGGCTTGAAG 3720
 QY 4424 GTGATATCCCTGGGCTGATTTGGGCTCTCACTGAAAGAAATGGCATCACCCCTTATCTTC 4483
 Db 3721 AAATACCATCCGCGTGGCATCTGTGAGAGAGTGCAGAGCTGGAATCATATGCTTCTGA 3780
 QY 4484 GCTGTGCTGCAATGTTTTCACATATTAATCTTGGGTTAATCCGCTGAGAACTGATATA 4543
 Db 3781 AGTGTCTGCTTATTTTTCATTTACTTAAATGAGCTTCCCTCCACCCGATTTCAAG 3840
 QY 4544 CCAATTCGAGAGAGAGATACAGTGCATCTGTAGTATCTATCTTACCTTACCAAAAT 4603
 Db 3841 T---TCTGGAACAAGCCATTTTGAACATTTATGATGATATCTTCTTCCATCAACCAAC 3897
 QY 4604 TGTTCCTCTCTTCCAGAAATTTTGGATATCTGTAAGGCCCTTGTCCAGAGAGGCTGTG 4663
 Db 3898 TCATTTGCTTTTCAAAATAATGATGATGATGATGATGATGATGATGATGATGATGATG 3957
 QY 4664 CAGATCTGCTTACTTAACCTTTGAAGCAAAAAACACCGTGTGAGATACCTTCAAGT 4723
 Db 3958 GTAAAGCAAGTAAAGATATCTAGAAGTGAAGAGATGCTATTAACATATCAAGATCA 4017
 QY 4724 AAAGAAATAGTTTGAAGCTTCTGATGATGATGATGATGATGATGATGATGATGATGATG 4783
 Db 4018 AATCTAAACAAATTAATTAACCTTCCAGAGATTTACAGCGCTCATTAATCAAGATCA 4077
 QY 4784 ATTTGAGTGGCCACGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 4843
 Db 4078 ATTTCTGCTGCGCCGAATCAGGTGTGATTAAGAGAGAGAGAGAGAGAGAGAGAGAG 4137
 QY 4844 GTGGGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4903
 Db 4138 GCGGATCTCTGTGCTGCCAGAGTACTGCTGCCAGACTGAGACTGAGAGAGAGAGAGATG 4197
 QY 4904 TTGAGCTTGCATTTTTCACGCACTTCACTGTGAGAGCCGAGCTGCTGATTTTCCATAAAA 4963
 Db 4198 TAGAGAGCTGCACACTCACACTCTCTGTGAGGCTGTGAGAGTGGGCTTCTGTAGAG 4257
 QY 4964 TCAGAGAAATGCGGAGTGTCTGTGTAAGGTAAGCAAGAGAGAGAGAGAGAGAGAGAGAG 5023
 Db 4258 TACGGAGATGCTAGTGTATTTTGAAGTGGGCAAAACCAAGGCTGTTTATTTATCTCTC 4317
 QY 5024 CTTAATGATGAATATGAG 5083
 Db 4318 CTTAATGATGAATATGAG 4377
 QY 5084 TATCTGAG 5143
 Db 4378 TATGAAAGAGAGAGATTAAGAGAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4437
 QY 5144 AGATTGCTAG 5195
 Db 4438 AAATGAG 4489

RESULT 8
 AAX03300
 ID AAX03300 standard; DNA; 1001 BP.

XX AAX03300;
AC
XX 25-MAR-1999 (first entry)
DT
XX Partial cDNA encoding a human ubiquitin-protein ligase, Ubri.
DE
XX Ubiquitin-protein ligase; Ubri; human; ubiquitinylation; degradation;
KW N-end rule pathway; stress-related muscle wasting; inhibitor; screen; ss.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH 1..999
FT CDS
FT /*tag= a
FT /product= Ubri
FT /note= "partial CDC"
XX
XX US5861312-A.
XX
XX 19-JAN-1999.
XX
XX 02-DEC-1997; 97US-0982956.
XX
XX 02-DEC-1997; 97US-0982956.
XX
XX (CALY) CALIFORNIA INST OF TECHNOLOGY.
XX
XX Kwon YT, Varshavsky A;
XX
XX WPI; 1999-130395/11.
DR P-PSDB; AAW84353.
XX
XX Mouse and human Ubri cDNA - useful for producing recombinant Ubri
PT polypeptides
XX
XX Claim 2; Columns 27-30; 18pp; English.
XX
XX The present sequence encodes a partial ubiquitin-protein ligase called
CC Ubri. The Ubri enzymes are involved in protein ubiquitinylation and
CC ultimate degradation through the N-end rule pathway and have been
CC linked to stress-related muscle wasting. Recombinant Ubri polypeptides
CC can be used to screen for inhibitors of muscle wasting when this is
CC associated with the N-end rule pathway.
XX
XX Sequence 1001 BP; 363 A; 186 C; 205 G; 247 T; 0 other;
SQ

Query Match 19.1%; Score 996.2; DB 20; Length 1001;
Best Local Similarity 99.7%; Pred. No. 8.9e-256;
Matches 998; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2059 ATGGATCCCAATAAGTCTTCTTACTGTGTTACTTCAGAGGTATGAAGTTCGCCGAGGCTTTT 2118
Db 1 ATGGATCCCAACAAGTCTTCTTACTGTGTTACTTCAGAGGTATGAAGTTCGCCGAGGCTTTT 60

Qy 2119 AACAGACCATATACAAAAGACGAGGATTTGATTAACAATAATATACACTAATAGAA 2178
Db 61 AACAGACCATATACAAAAGACGAGGATTTGATTAACAATAATATACACTAATAGAA 120

Qy 2179 GAAATGCTCAGGTCCTCATCTATATTTGGGGTGGAGCTTATGACCTGGAGTGGGAAT 2238
Db 121 GAAATGCTCAGGTCCTCATCTATATTTGGGGTGGAGCTTATGACCTGGAGTGGGAAT 180

Qy 2239 GTGACCAAAAGAGGTCACATCAGAGAAATCATCTTGTGCTTGAACCCCATG 2298
Db 181 GTGACCAAAAGAGGTCACATCAGAGAAATCATCTTGTGCTTGAACCCCATG 240

Qy 2299 CCACACAGTGCATTTGCCAAAATTTACCTGAGAAATGAATGAACCTGGCTTAGAG 2358
Db 241 CCACACAGTGCATTTGCCAAAATTTACCTGAGAAATGAATGAACCTGGCTTAGAG 300

Qy 2359 AATGTCATAACAAAGTGGCCACATTTAAGAAACAGGTGATCAGGCCATGGAGTTTAT 2418
XX

Db 301 AATGTCATAAACAAAGTGGCCACATTTAAGAAACACAGGTGTATCAGGCCATGGAGTTTAT 360
Qy 2419 GAACATAAAGATGAATCAATGAAGACTTCAATATGTACTTTTATCATTTACTCCAAACC 2478
Db 361 GAACATAAAGATGAATCAATGAAGACTTCAATATGTACTTTTATCATTTACTCCAAACC 420
Qy 2479 CAGCATACAGGCTCAACATATGCAGAGAAAGAGGAGAAACAAACAAACAAAGATGAA 2538
Db 421 CAGCATACAGGCTCAACATATGCAGAGAAAGAGGAGAAACAAACAAACAAAGATGAA 480
Qy 2539 GCATTGCGGCCACACACCTCTCTGAATTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCT 2598
Db 481 GCATTGCGGCCACACACCTCTCTGAATTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCT 540
Qy 2599 CTCAACTGTGATATCATGATGTACATTTCTCAGAGACCGTATTTGAGCGGCAATAGACACA 2658
Db 541 CTCAACTGTGATATCATGATGTACATTTCTCAGAGACCGTATTTGAGCGGCAATAGACACA 600
Qy 2659 GATTCTAACTTGTGGACCGAAGGATGCTCCAAATGGCTTTTCATATTTCTGGCATTTGGGT 2718
Db 601 GATTCTAACTTGTGGACCGAAGGATGCTCCAAATGGCTTTTCATATTTCTGGCATTTGGGT 660
Qy 2719 TTACTAGAGAGAGCAACAGCTTCAAAAAGCTCTCTGAAGAGAAAGTAAACATTTGACTTT 2778
Db 661 TTACTAGAGAGAGCAACAGCTTCAAAAAGCTCTCTGAAGAGAAAGTAAACATTTGACTTT 720
Qy 2779 TATCATAAGGCTTCAAGATTTGGAAAGTTCAGCCATGAATATACAAATGCTTTTGGAAAAA 2838
Db 721 TATCATAAGGCTTCAAGATTTGGAAAGTTCAGCCATGAATATACAAATGCTTTTGGAAAAA 780
Qy 2839 CTCAAAGGNAATCCCCAGTTAGAGCCAGAGGACATGATACGTGGTACTTCTCAGATG 2898
Db 781 CTCAAAGGNAATCCCCAGTTAGAGCCAGAGGACATGATACGTGGTACTTCTCAGATG 840
Qy 2899 TTTGACACAGTGAAGCGATTAAAGAGAAAAATCTTTTAAATTTAGCAACACCATCAGGA 2958
Db 841 TTTGACACAGTGAAGCGATTAAAGAGAAAAATCTTTTAAATTTAGCAACACCATCAGGA 900
Qy 2959 TCGGAATCTATTAAAGATGATGAGATTACTCATGATATAAGAAAAAGCAAGCAAGAAAAAGA 3018
Db 901 TCGGAATCTATTAAAGATGATGAGATTACTCATGATATAAGAAAAAGCAAGCAAGAAAAAGA 960
Qy 3019 AAAGCTGAAGCTGTAGGCTATGCTCCAGAGAGATCATGGC 3059
Db 961 AAAGCTGAAGCTGTAGGCTTCTATCGCCAGAGATCATGGC 1001

RESULT 9
AAC86934
ID AAC86934 standard; cDNA; 1001 BP.
XX
AC AAC86934;
XX
DT 02-APR-2001 (first entry)
XX
DE Nucleotide sequence of a partial human Ubri protein.
KW Ubri; E3-type protein; ubiquitin system; ubiquitin-protein ligase;
KW N-end rule pathway; intracellular pathogen; Lysteria monocytogenes;
KW Yersinia enterocolitica; muscle wasting; infection; ss.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH 1..999
FT CDS
FT /*tag= a
FT /product= "Ubri"
FT /note= "partial sequence"
XX
XX US6159732-A.
XX
PD 12-DEC-2000.
XX

PF 11-JAN-1999; 99US-0228317.
PR 02-DEC-1997; 97US-0982956.
XX (CALY) CALIFORNIA INST OF TECHNOLOGY.
PA Kwon YT, Varshavsky A;
PI WPI: 2001-090278/10.
XX P-PSDB; AAB31163.
DR
XX Inhibiting the N-end rule pathway in mammalian cells for treating
PT infections and various diseases associated with muscle tissue wasting,
PT by inhibiting the expression of Ubr1 gene
PS
PS Claim 4; Column 27-30; 18pp; English.
CC The present sequence encodes a partial Ubr1 enzyme. Ubr1 is an E3-type
CC protein of the ubiquitin system. Specifically, it is a ubiquitin-protein
CC ligase. The enzyme is specific for destabilising residues exposed at
CC the N-terminus of protein substrates. Inhibition of the expression of
CC Ubr1 gene in a cell results in inhibition of the N-end rule pathway.
CC The method is used for treatment of mammalian cells infected with an
CC intracellular pathogen, e.g., *Lysteria monocytogenes* or *Yersinia*
CC enterocolitica. Inhibition of N-end rule pathway is also useful for
CC treating various diseases associated with wasting of muscle tissue and
CC infections.
CC
SQ Sequence 1001 BP; 363 A; 186 C; 205 G; 247 T; 0 other;

Query Match 19.1%; Score 996.2; DB 22; Length 1001;
Best Local Similarity 99.7%; Pred. No. 8.9e-256;
Matches 998; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2059 ATGATCCCAATTAAGTCTTGTACTGCTGACTTCAGAGGTATGAACCTGGCGGCTTT 2118
DB 1 ATGATCCCAATTAAGTCTTGTACTGCTGACTTCAGAGGTATGAACCTGGCGGCTTT 60
QY 2119 AACAGACCATATCTACAAAGAGCAGATTGATTAAACAATATATACATATATAGAA 2178
DB 61 AACAGACCATATCTACAAAGAGCAGATTGATTAAACAATATATACATATATAGAA 120
QY 2179 GAAATGCTTCAGGTCTCTATATATGTTGGGTGAGCCGTATATGCTGAGTGGCAAT 2238
DB 121 GAAATGCTTCAGGTCTCTATATATGTTGGGTGAGCCGTATATGCTGAGTGGCAAT 180
QY 2239 GTGACCAAGAAGAGGTGCATGAGAAATCATTCCTGCTTGCATTTGAAACCCATG 2298
DB 181 GTGACCAAGAAGAGGTGCATGAGAAATCATTCCTGCTTGCATTTGAAACCCATG 240
QY 2299 CCACACAGTGCATTCGCAAAATTTTACCTGAGAAATGAATAATGAATGAGCTTTAGAG 2358
DB 241 CCACACAGTGCATTCGCAAAATTTTACCTGAGAAATGAATAATGAATGAGCTTTAGAG 300
QY 2359 AATGTCATTAACAAGGTGCGCCACATTTTAAAGAACAGGTGATATCAGGCCATGAGTTTAT 2418
DB 301 AATGTCATTAACAAGGTGCGCCACATTTTAAAGAACAGGTGATATCAGGCCATGAGTTTAT 360
QY 2419 GAACATAAAGATGATTCATCTGAAGACTCATATGATCTTTATATCTATACCTCAAAAC 2478
DB 361 GAACATAAAGATGATTCATCTGAAGACTCATATGATCTTTATATCTATACCTCAAAAC 420
QY 2479 CAGCATAGCAAGGTGATCATATGAGAAAGAGAGAAACAAACAAAGATGAGAA 2538
DB 421 CAGCATAGCAAGGTGATCATATGAGAAAGAGAGAGAAACAAACAAAGATGAGAA 480
QY 2539 GCATTGCGCGCAACCAACCTCTGAAATTCGCCCTTTCAGCAAAAGTGAATACCTT 2598
DB 481 GCATTGCGCGCAACCAACCTCTGAAATTCGCCCTTTCAGCAAAAGTGAATACCTT 540
QY 2599 CTCAACTGTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2658
DB 541 CTCAACTGTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600

QY 2659 GATTCTAACYTTGTGACCGAAGAGATGCTCCAAATGCGTTTCATATTCGTGCAATGGGT 2718
DB 601 GATTCTAACYTTGTGACCGAAGAGATGCTCCAAATGCGTTTCATATTCGTGCAATGGGT 660
QY 2719 TTACTAGAGAGCAACACCTTCAAAAGCTCTGAGAGAGAGATTAACATTTGACTTT 2778
DB 661 TTACTAGAGAGCAACACCTTCAAAAGCTCTGAGAGAGAGATTAACATTTGACTTT 720
QY 2779 TATCATTAAGGCTTCAGATTTGGGAAGTTCCAGCATGAAATATACAAATGCTTTGGAAAA 2838
DB 721 TATCATTAAGGCTTCAGATTTGGGAAGTTCCAGCATGAAATATACAAATGCTTTGGAAAA 780
QY 2839 CTCAGAGGAATTCGCCAGTTTGAAGGCCAGAGACATGATAGCTGTGATCTTCAAGATG 2898
DB 781 CTCAGAGGAATTCGCCAGTTTGAAGGCCAGAGACATGATAGCTGTGATCTTCAAGATG 840
QY 2899 TTGGACACAGAGAGCATTTAGAGAAAAATCTTTAATTTGTGCAACCATCAGAGA 2958
DB 841 TTGGACACAGAGAGCATTTAGAGAAAAATCTTTAATTTGTGCAACCATCAGAGA 900
QY 2959 TCGGAATCTATTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3018
DB 901 TCGGAATCTATTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
QY 3019 AAGCTGAACCTGCTAGCTACATCGCCAGAGATCATGCG 3059
DB 961 AAGCTGAACCTGCTAGCTACATCGCCAGAGATCATGCG 1001

RESULT 10
AAK35730
ID AAK35730 standard; cDNA: 3327 BP.
XX
AC AAK35730;
XX
XX 09-JUL-1999 (first entry)
DE
XX cDNA encoding a protein identified by the signal sequence trap method.
XX
XX Signal sequence trap method; SST method; immunisation; inhibition;
XX infection; allergy; cancer; regulation; tissue formation; tissue repair;
XX cytokine activity; blood coagulation regulation; agonist; antagonist;
XX metabolic disorder; hormonal disorder; immune disorder;
XX severe combined immunodeficiency; SCID; AIDS; thrombosis; cancer;
XX wound; ss.
OS Homo sapiens.
XX
XX W09918126-A1.
XX
XX 15-APR-1999.
XX
XX 06-OCT-1998; 98WO-JP04514.
XX
XX 07-OCT-1997; 97JP-0274674.
XX
XX (ONOV) ONO PHARM CO LTD.
XX
XX Fukushima D, Shibayama S, Tada H;
XX
XX WPI: 1999-277254/23.
XX
XX P-PSDB; AAY02376.
XX
XX Polypeptides identified by the signal sequence trap method from a
XX human cDNA library
XX
XX Claim 5; Page 167-169; 281pp; Japanese.
XX
XX AAK35694-X35747 represent cDNA sequences that encode novel polypeptides
XX (AAY02358-84) which are identified from a human placental cDNA library
XX by the signal sequence trap (SST) method. The polypeptides have a

CC	broad range of physiological activity, including immunisation against
CC	and inhibition of infections, allergies and cancer; regulation of tissue
CC	formation and repair; activin/inhibin activity; chemokine/cytokine
CC	activity; blood coagulation regulation; and receptor/ligand agonist
CC	or antagonist activity. The polypeptides can be used for prevention
CC	and treatment of disorders including infections by bacteria, yeasts and
CC	viruses (including HIV) and protozoa; metabolic and hormonal disorders;
CC	immune disorders (including severe combined immunodeficiency (SCID)
CC	and AIDS; thrombosis; cancer; and traumatic or surgical wounds.
XX	
SQ	Sequence 3327 BP; 1033 A; 666 C; 739 G; 889 T; 0 other;
	Query Match 15.6%; Score 812; DB 20; Length 3327;
	Best Local Similarity 55.3%; Pred. No. 3.3e-206;
	Matches 1855; Conservative 0; Mismatches 1410; Indels 87; Gaps 11;
QY	1892 TGGAAATATCCTTACGTTGTCTCGTGTTGGTTGCCAGGTTGTTCGTGAGATGTGGCAA 1951
DB	
DB	8 TAGAACACCCTCTTAGATGTCCTTTGTTGTGTGCCCAAGTACATGTCGGGAATGTGGAA 67
QY	1952 GAAATGACATGCTCTCTATTAGCGAGGTGTTTTATTACCAAGATGTTTAAGTCGAGAAG 2011
DB	
DB	68 GAATGGGTCTCTCTAGTAGTAACACAGATTATTACTACCAATAATGTGAATGCAGACGTG 127
QY	2012 AAATGTATGAAGAATATCATCATGCTTCAGATTTGGTGCATCTTTAATGGATCCCAATA 2071
DB	
DB	128 AGATGTTTGACAAGGATGTAGTAATGCTTCAGACAGGTGTCTCCATGATSGATCCAATC 187
QY	2072 AGTTCCTGTTACTGGTACTTCAGAGGTATGAACTTGGCGAGGCTTTTAAACAAGCCATAT 2131
DB	
DB	188 ATTTCCTGATGCATCATCTCAGCGCTTTTGAACCTTTATCAGATTTTTCAGTACTCCAAGACT 247
QY	2132 CTACAAAA----- -CACGAGGATTTGATTAAACAATATAAATACAC 2170
DB	
DB	248 ATGAAAAGAATTTAGTCTGAGATTACCCATAGGATGTTGTTTCAGCAACAATACTC 307
QY	2171 TAATAGAAGAAATGCTTCAGGTCTCATCTATATTTGTGGGTGAGCGTTATGTACCTGGAG 2230
DB	
DB	308 TAATAGAAGAAATGCTATACCTCATTTATATGCTTGTGGAGAGAGATTTAGTCTCGGAG 367
QY	2231 TGGAAATGTGACCAAGAAGAGGTACAATGAGAGAAATCATTTCACTTCTCTTTGCATTG 2290
DB	
DB	368 TTGACAGGTAAATGCTTACAGATGAATCAAGCGAGAGATTTCCATCACTTCAGTATCA 427
QY	2291 AACCCATGCCACAGTGCATTTGCCAAAAATTTTACCTGAGATGAAATAATCAAACTG 2350
DB	
DB	428 AGCCTATGGCTCATAGTGAATTTGGTAAAGTCTTTTACCTGAAGATGAGAACAGAGAGACT 487
QY	2351 GCATTAGAGAAATGTCATAAACAAAGTGCCACATTTTAAGAACACGAGTGTATCAGGCCATG 2410
DB	
DB	488 GCATGGAGAGTGAATCGAAGCAGTTGCCCATTTCAAGAAACCTGGATTACAGGACGAG 547
QY	2411 GAGTTTATGAATAAAGATGAATCACTGAAAGACTTCAATATGTACTTTTATCATTTACT 2470
DB	
DB	548 GCATGTATGAATGAACACCAAGATGTGCCAAAGAGTTCAACTTGTATTTCTATCACTTTT 607
QY	2471 CCAAAACCCAGCATAGCAAGGCTGAACATATGCAAGAAGAAAGAGGAAACACAGAAACA 2530
DB	
DB	608 CAAGGGCAGAACAGTCCGAAGCAGAAAGCGCAACGCAAAATTTGAAAGAGCAAAATAGAG 667
QY	2531 AAGATGAAGCATTTGCCGCCACCACCACTCTCTGAATTTCTGCCCTGCTTTTCAGCAAAAGTGA 2590
DB	
DB	668 AAGATACAGCACTCCCACCTCCGGTGTTCCTCCATTTCTGCCCTCTGTTTGCAGAGCCTGG 727
QY	2591 TTAACCTTCTCAACTGTGATATCATGTATGATGTACATTTCTCAGGACCGGATTTTGACGGGGCAA 2650
DB	
DB	728 TTAACATTTTGCAGTCAGATGTCATGTTGTGCATCATGCGGAACAATTTCTGCAATGGGCTG 787
QY	2651 TAGACACAGATTTAACTTGTGGACCGAAGGATGCTCCAAATATGGCTTTTCATATTTCTCG 2710
DB	
DB	788 TGGAAACATAATGGATATGCGCTGGTTCAGATGCTCCATGCTGCAAGGGGTGTACATTTAATTG 847
QY	2711 CATTTGGTGTACTAGAGAAGCAACAGCTTCAAAAAGCTCTCTGGAAG----AAGAAGTAA 2767

Db 1907 AGCTCCCTGAGAGGCTTCAGGCTGATTTTCGCTTAAGATCCCTTATCTGAGACATTA 1966
 QY 3827 AGGAATGGTATTTCTCTTGGCCACAACATTTATAGATTTGATGAAAGTCCACCTG 3886
 Db 1967 AAGAAATGCTTAAGAGATTTGAACTGTAACCAAGGTTGGGATTAAGGTTTCATGCCA 2026
 QY 3887 ATGAAAGGATCTCGAGTCCCATCTGCTGAGACACCTGCGCTTTCATATCCAGG 3946
 Db 2027 ATGAAAGGATCTCGGTTTCCATATATGTTGGGGTAGCTGCGGCTACACATCCAAA 2086
 QY 3947 CAATGAAAATCTATTTGGAGTGAAGAAACCTGTTTGGAGACCTTCAAAATAGGC 4006
 Db 2087 GATGAGAAAGAAATTTTGAAGTGAAGAAATTAACCATTTGTTGCTTTACCTTGACAC 2146
 QY 4007 AGCATATATGCTGTAAGCATTAATGACATTTGACATTCACAGAGATTAACCTGCTC 4066
 Db 2147 TGGATGACGTCTTAAAGCATTTGACAGATTTGGCCGACACACCTGACAGTGGCAAG 2206
 QY 4067 AGGCTCTGATACAGAAACATCTGCTGCTCTATCAGTTGTTCTTCAATCAAT 4126
 Db 2207 TTTGAGTGTGACAGAGACATTTTGTAACTTTTTCATCAGCTGCTCAATGACAGCC 2266
 QY 4127 CAGAGATACACCATGCTCTGCTATATGATCTGTTTATGTTGGGCTGCTGTG 4186
 Db 2267 ATGAGGAACCTTCTGATATGATATGATATTAACCATTTGTTGCTTTACCTTGACAC 2326
 QY 4187 TAGCATTCCTCATCTGATTTGAGATGACCTGTTGATCTGACGCTTCTTCACTAGT 4246
 Db 2327 TTGCATTTCTGCTGCTGACGTG-----TAGGATTTTTCAGGATACGCC 2371
 QY 4247 CTTCCTATTAACCACTTTATCTCTTCCATTTGATACATGACATGACATGCTTACATAC 4306
 Db 2372 TTGGGACATGGAACCTTCCATTTTCCATCTGTTACATGACACATCATCATCATCT 2431
 QY 4307 TACTTACAGTACAC---AGGCTACCCCTGCTCAGGTTCAAGAAACAGTGAAGAG 4363
 Db 2432 TACTTACCTCATGTACAGAAATGATGATGATTAACAAATCCCTCTGTAAGAG 2491
 QY 4364 CTGATTCGCGATCTCTTCTTCTTGGAGAAATTTCTCAATATACAAAGTGGCTCATGG 4423
 Db 2492 AATGACGAGTTCTGCTTGTATTAACACCTTCACAGATATACGGAAGTGCCTTAAAG 2551
 QY 4424 GTGATATTCCTGCTGCTGATTTTGGGCTCACTGAAGATGACATCCCTTATCTTC 4483
 Db 2552 AAATACCAATCCGCTGCTGCTGCTGAGAGTGTGACGATGGAATCATGCTTCTCTGA 2611
 QY 4484 GCTGCTGCTGCTGCTTTCCTACTATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4543
 Db 2612 AGTGTCTGCTTATTTTCTTCACTTAATGGAATGCTTCTCCACCCGACATTCAG 2671
 QY 4544 CCAATTCGAGAGAGAGATGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4603
 Db 2672 T---TCTGGAACAAGCCATTTTGAACATTTATGATGATTTCTTCTCCACCAACACC 2728
 QY 4604 TGTTCCTGCTCTCTCAAGAAATTTGGGATCTGTAAGGCTTCTGCTCAAGAGGCTGTG 4663
 Db 2729 TCATTTGCTTTTTCAGAAATATGAGATATGATTAATCAGATTTGAAGTGTGCTG 2788
 QY 4664 CAGATCTGCTTACTAACTGTTGAGCAAAAACACCGTGTGCTGCTGCTGCTGCTGCT 4723
 Db 2789 GTAAAGTGAAGTAAAGATATCTAGAAGTGAAGAGATCTATTAATATCTCAAGAG 2848
 QY 4724 AAAGAAATGTTGATGAGCTTCTGATGATCTATGCTGCTGCTGCTGCTGCTGCTGCT 4783
 Db 2849 AATCTAAACAATTAATTAACCTTCAAGAGATTTACAGAGCTCATTAATCAAGCATCA 2908
 QY 4784 ATTTGAGGCTCCAGGCTGCTGAGATGATGAGCAAGATCTGCTGCTGCTGCTGCTGCT 4843
 Db 2909 ATTTTCTGCTCCGAAATACAGTGTGATTAAGACAGAGCCCAACTGTGCTGCTGCTG 2968
 QY 4844 GTGGGCTATATCTATGTTCTGAAACATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4903
 Db 2969 GCGGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3028

QY 4904 TTGAGCTTGCAATTTTTCACGCACTTCACTGTGAGCGGAGTCTGATTTTCTAATAA 4963
 Db 3029 TAGAGCTGCAAGCACTACACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3088
 QY 4964 TCAGAGAAATCCGAGTGTCTGTTGAAGTAAAGCCAGAGGCTGTGCTTATCCAGTCC 5023
 Db 3089 TACGGGAATGTCAGTCTATTTTAACTGCTGCAAAACCAAGGCTTTTATTTCTCTC 3148
 QY 5024 CTTCCTGATGATATGAGAAACAGACCTGCTGCAAGAGGAGGCAACCCCTTCAT 5083
 Db 3149 CTTCCTGATGATATGAGAAACAGACCTGCTGCAAGAGGAGGCAACCCCTTCAT 3208
 QY 5084 TATCTGAGCGGCTATCGAAGCTTCACTTTGCTGCAACACACCTGCTATATAGAAG 5143
 Db 3209 TATGCAAAAGCGATTTCAAGAAATTCAGAAAGCTGCTGCTGCTGCTGCTGCTGCTG 3268
 QY 5144 AGATGCTGAGAGCCCAAGACATATCATGATTTATTTGATTCACATGCGCA 5195
 Db 3269 AAATGGAATGTCACAGGAAGCCATGACACGCTGTTGCTGCTGCTGCTGCTGCTG 3320

RESULT 11
 AAX35731
 ID AAX35731 standard; cDNA; 3502 bp.
 XX
 AC AAX35731:
 XX
 DT 09-JUL-1999 (first entry)
 XX
 DE cDNA encoding a protein identified by the signal sequence trap method.
 XX
 KW Signal sequence trap method; SST method; immunisation; inhibition;
 KW infection; allergy; cancer; regulation; tissue formation; tissue repair;
 KW activin activity; inhibin activity; chemokine activity;
 KW cytokine activity; blood coagulation regulation; agonist; antagonist;
 KW metabolic disorder; hormonal disorder; immune disorder;
 KW severe combined immunodeficiency; SCID; AIDS; thrombosis; cancer;
 KW wound; ss.
 XX
 OS Homo sapiens.
 XX
 PN W09918126-A1.
 XX
 PD 15-APR-1999.
 XX
 PF 06-OCT-1998; 98WO-JP04514.
 XX
 PR 07-OCT-1997; 97JP-0274674.
 XX
 PA (ONOV) ONO PHARM CO LTD.
 XX
 PI Fukushima D, Shibayama S, Tada H;
 XX
 WP1: 1999-277254/23.
 DR P-PSDB: AAY02376.
 XX
 PT Polypeptides identified by the signal sequence trap method from a
 PT human cDNA library
 XX
 PS Claim 4: Page 170-178; 281pp; Japanese.
 XX
 CC AAX35694-X35747 represent cDNA sequences that encode novel polypeptides
 CC (AAY02356-84) which are identified from a human placental cDNA library
 CC by the signal sequence trap (SST) method. The polypeptides have a
 CC broad range of physiological activity, including immunisation against
 CC and inhibition of infections, allergies and cancer; regulation of tissue
 CC formation and repair; activin/inhibin activity; chemokine/cytokine
 CC activity; blood coagulation regulation; and receptor/ligand agonist
 CC or antagonist activity. The polypeptides can be used for prevention
 CC and treatment of disorders including infections by bacteria, yeasts and
 CC viruses (including HIV) and protozoa; metabolic and hormonal disorders;
 CC immune disorders (including severe combined immunodeficiency (SCID))

CC and AIDS; thrombosis; cancer; and traumatic or surgical wounds.

Sequence 3502 BP; 1093 A; 690 C; 760 G; 959 T; 0 other;
XX
SQ

Query Match	15.6%	Score 812;	DB 20;	Length 3502;
Best Local Similarity	55.3%	Pred. NO. 3.4e-206;		
Matches 1855;	Conservative	0;	Mismatches 1410;	Indels 87

Qy	1992	TGGAATATCCTTTACGTTGTCTGGTGTGTTGTTGCCACAGGTTGTGCTGAGATGTGCGCAA	1955
Db	63	TAGAACACCCTCTTAGATGCTTGTGTCGTGTCGCCAAGTACATGCCGGAATGTGGAGAA	122
Qy	1952	GAATGGACTGCTCTCTTATTAGCCAGGTGTTTATTACCAAGATGTTAAGTCGAGAGAAG	2011
Db	123	GAATGGGTCTCTCTAGTAAACACAGATTTATTACTACCATAATGTGAAATGCAGACGTG	182
Qy	2012	AAATGTATGATAAAGATATCATCATGCTTCAGATTTGGTGCATCTTTAATGGATCCCAATA	2071
Db	183	AGATGTTTGACAGGATGTAGTAATGCTTCAGACAGGTGCTCCATGATGATCCCAATC	242
Qy	2072	AGTTCTTGTACTGGGTACTCTCAGAGGTATGAACCTGCGGAGGCTTTTAAACAAGACATAT	2131
Db	243	ATTTCTCTGATCATGCTCAGCGCTTTGAACTTTATCAGATTTTTCAGTACTCCAGACT	302
Qy	2132	CTACAAA-----CACCAGGATTTGATTAAACAATATAATACAC	2170
Db	303	ATGGAATAAGATTAGTTCTGAGATTAACCATAAGGATGTGTTTCAGCAGAACAATACTC	362
Qy	2171	TAATPAGAAGAAATGCTTCAGGTCTCATCTATATTGTGGTGAGCGGTTATGTACCTGGAG	2230
Db	363	TAATAGAAGAAATGCTATACCTCATTAATGCTTGTGGAGAGAGATTTAGTCTCGGAG	422
Qy	2231	TGGGAATGTGCAAAAGAGAGAGGTGACAATGAGAGAAATCATTCAGTCTGCTTTGCAATG	2290
Db	423	TTGACACAGGTAATGCTACAGATGAATCAAGCGAGAGATTATCCATCAGTTGAGTATCA	482
Qy	2291	AACCCATGCCACACAGTGCCATTGCCAAAAATTTACCTGAGATGAAAAATAAGTAAACTG	2350
Db	483	AGCCTATGGCTCATGTGTAATTTGGTAAGTCTTTTACCTGAAGATGAGAACAAGAGACTG	542
Qy	2351	GCTTAGAGAATGTCATAAACAAGTGGCCACATTTAAGAAACCAAGGTGTATCAGGCCATG	2410
Db	543	GCATGGAGAGTGAATCGAAGCAGTGTGCCCATTTTCAGAAACCTGGATTTAACAGGACGAG	602
Qy	2411	GAGTTTATGAATAAAAGATGAATCACTGAAAGACTTCATATGTACTTTTATCATTTACT	2470
Db	603	GCATGTATGAAGTGAACCAAGATGTGCCAAAGAGTTCAACTGTGTTTCTTCTACATTTT	662
Qy	2471	CCAAAACCCAGCATAGCAAGGCTGAACATATGCAGAAGAAAAGAGAAAAACAAGAAAAACA	2530
Db	663	CAAGGCAGACAGTCCAGGCAGAGAGCGCAACCGGAATTCGAAAGACAAATAGAG	722
Qy	2531	AAGATGAAGCAATGCCGCCACCAACCTCCTGAAATTCGCGCTGCTTTTCAGCAAAAGTGA	2590
Db	723	AAGATACGCACTCCCACCTCCGGTGTGCTCCTCATTCGCGCTCTGTTGCAAGCGCTGG	782
Qy	2591	TTAACCCTCAACTGTGATATCATGATGTACATCTTCAGGACCGTATTTCAGCGGCGAA	2650
Db	783	TTAACAATTTGCAGTGCAGATGTCAATGTGTGCATCATGCGGAACAATTTCTGCAATGGCTG	842
Qy	2651	TAGACACAGATTCTTAACCTGTGTGGACCGAAGGATGCTCCAAATGGCTTTTCATATTCTGG	2710
Db	843	TGGAACATATGGATATGCTCTGGTTCAGATCCATGCTGCAAGGGTGTACATTTAATTG	902
Qy	2711	CATTGGGTTTACTAGAAGAGAAAGCAACAGCTTCAAAAAGCTCCTTGAAG---AGAAGTAA	2767
Db	903	GCATGGCACTACAAGAGAAAAACAATTTAGAGAATGTCAAGGAAAGCATGTAGTAA	962
Qy	2768	CATTTGACTTTTATCATTAAGCTTTCAAGATTGGGAAGTTTCAGCCATGAAT-----A	2818
Db	963	CATTTACTCTCCTCAGAAGATATCAAAACCTGTGTGAAGCGCCAAAATAATTTCTCCTAGCA	1022
Qy	2819	TACAAATGCTTTTGGAAAACTCAAGGAAATTTCCCAAGTTAGAGGGCCAGAAGGACATGA	2878

Db	1023	TACTAGCTATGCTGGAAACACTACAAAAATGCTCCCTACCTAGAAGTCCCAAAAGACATGA	1082
Qy	2879	TAACGCTGGATACTTCAGATCTGTTGCACACAGCTGAAGCGATTAAAGAGAAAAATCTGTTTAA	2938
Db	1083	TTCCGCTGGATATTGAAGACTTTTAATGCTGTTAAAGATGAGGGAGA-----GTTCAC	1136
Qy	2939	TTGTAGCAACCACATCAGGATCGGAATCTATTAAAGATGATGAGTTACTTCATGATAAAG	2998
Db	1137	CTACCAGTCCGCTGGCAGAGACAGAAGAACCATTAATGGAAGAGAGTTCAAGGACAAAG	1196
Qy	2999	AAAAAGCAGAAACGAAAAAGAAAAGCTGAAGCTGCTAGGCTACATCGCCAGAAAGCATGCG	3058
Db	1197	ACAAAGCTGAGAGGAAGAAAAAGCAGAGATTGCCAGACTGCCAGAGAAAAGATCATGCG	1256
Qy	3059	CTCAGATGCTCGCCTTACAGAAAAAATTCATTGAAACTCATAAACTCATGATGACAAATA	3118
Db	1257	CTCAGATGCTGGAATGACGGGCAATTTATGATGAAAACAAGAAGACTCTTTCAGCAGA	1316
Qy	3119	CATCAGAAATGCCCTGGGAAAGAAGATTCCCATTTATGGAGGAGAGAGACACCCAGCAGTCA	3178
Db	1317	CATTAGAACGCGATGCGCTCAACCTCTGCTGTTCTTGATCA-----TAGCCCTGTGCGTT	1370
Qy	3179	GTGACTACTCTAGAAATTTGCTTTGGGTCCCTAAACGGGGTCCATCTGTTACTTGAAGAGAGG	3238
Db	1371	CAGATATGACACTTTACAGCAGCTGGGCGCCGCACAAACTCAGGTTCTCTGAAACAAGACAAT	1430
Qy	3239	TGCTGAGCTGCATCCTTTGGCCAAAGAACAGAGAGGTGAAATAGAAAAATTAATGCCATGG	3298
Db	1431	TCGTTTACATGATATATGTGTCAAGGAGGACAAGATTTAAAGTGGAAAGCAGGGCAGTGG	1490
Qy	3299	TATTATCGGCCTGTGCCAGAAATCTACTGCTTTAACCCAGCAGACAGGGGAAAAACCATAG	3358
Db	1491	TCTTGGCAGCATTTGTTCAGAGATCAACTGTATTATCAAAAACAGAGAATTAATTTATTC	1550
Qy	3359	AACCTCTCAGGAAGACCCCTAGACCCACTTTTCATGATCCAGACTTGGCATATGGAACCTT	3418
Db	1551	AAGATCCAGAAAAATAT--GATCCATTATTCATGACCCCTGATCTGCTGTGTGGAAACAC	1607
Qy	3419	ATACAGGAAGCTGTGGTCATCTAATGACGACAGTGTCTGCGCAGAGATATTTTGAAGCTG	3478
Db	1608	ACACTAGTAGTGTGGGCACATATGATGACCCCATTTGTTGGCAAGGATATTTGATTCGG	1667
Qy	3479	TACAGCTGAGCTCTCAGCAGCGCATTCATGTTGACCTTTT-----TGACTTGG	3526
Db	1668	TTCAAGCTAAAGAACAGCGAAGGCAACAGAGATTTACGCTTACATACAGCTATGATGTAG	1727
Qy	3527	AAAGTGGAGAAATATCTTTGGCCCTCTTTGCAAAATCTGTGCAATCTGTGATCCCCATTA	3586
Db	1728	AAAAAGGAGAAATCTTTTGGCCCTTTGTGAATGCTTGAATGCTTGAATGCTTGAATGCTTGA	1787
Qy	3587	TTCTCTTTCACCTCAAAAGATAAACAGTGTAGATGCGATGCTTCTTGCTCAACTTTTGA	3646
Db	1788	TGCTTT-----TTCCAAGAAATATTTTTAACACAGGTTAAATTTTCAGACCAACCAA	1841
Qy	3647	CCCTGGCAGCGTGGATACAGACTGTTCTGGCCAGAATATCAGGTTATAATATAAGACATG	3706
Db	1842	ATCTCACTCACTGGATTAGACAAATATCTCAGCAAAATAAAAGCATTACAGTTTCTTTAGGA	1901
Qy	3707	CTAAGGAGAAACCCCAATTCCTATTTCTTTTAATCAAGGAATGGGAGATTCCTACTTTGG	3766
Db	1902	AAGAGAAAGTACTCCTAATAATGCCTCTACAAAAGAAATTCAGAAAAATGTGGATGAATTAC	1961
Qy	3767	AGTTCCATTCCTCCTAGTGTGGCGTTGAGTCTTCGATTAATAATTTCAAAATAGCATCA	3826
Db	1962	AGCTCCCTGAAGGGTTCAGGCGCTGATTTTGTGCTTAAGATCCCTTATTTCTGAGAGATAA	2021
Qy	3827	AGGAAATGGTTATTTCTTTTGGCCACAACAATTTATAGAATTTGGAATGAAAGTGCCACCTG	3886
Db	2022	AAGAAATGCTAAGCACAATTTGGAACTGCTACCTACAGGTGGGAGTAAGGTTTCATCCCA	2081
Qy	3887	ATGAAGGAGTCTCTCGAGTCCCCCTGCTGACCTGGAGCACCTGCGCTTTTCACTATCCAGG	3946

Db 2082 ATGAAGAGATCCGCTGTTCCCATTAATGTGTTGGGGTAGCTGCGGTAACACATCCAAA 2141
 QY 3947 CAATTGAAATATCTATTTGGAGATGAAGAAACCTCTGTTGGAGACTTCAAAATAGGC 4006
 Db 2142 GCATGAAAGAAATTTGAGTGAATGAAGATTAACCATTTGTTGGTCTTATACCTTGCAGAC 2201
 QY 4007 AGCATTAATGTCCTGAAGCATTAATGATGAGTTGACAGTTGCAGAGAGATTACCTGTCAC 4066
 Db 2202 TGGATGACTGCTTATAGTCAATTTGAGAGATTTGGCCGACACACTGAGAGTGCATCAG 2261
 QY 4067 AGTCTGATACAGAAACATCTGTTGCTCTTATACATGTTCTTCTTAACATTAAT 4126
 Db 2262 TTTCAGTGGTCAAGACATTTTGTAACTTTTGCATCACTGGTGCCTAATGACAGCC 2321
 QY 4127 CAGAAGATACACACATGCTCTGCTATAGATCTGTTCACTGTTGGTGGTCTGCT 4186
 Db 2322 ATGAGAACTTCCATGATATATAGATATGACATTTTCATTTATGTTGGTGGTGGTGC 2381
 QY 4187 TAGCATTCACATCCCTGATATGGAGTGAACCCGTTGATCTGACGCTTCTCAGTTAGTT 4246
 Db 2382 TTGCATTTCTGCGTTGCACTG-----TCAGGATTTTTCAGGGATCAGCC 2426
 QY 4247 CTTCCTATACACCTTTATCTCTTCCATTTGATCACCATGACACATGCTTACAGATAC 4306
 Db 2427 TTGGCACTGAGACCTTCACATTTCCATCTGTTACTATAGGACACATCATACAGATCT 2486
 QY 4307 TACTTACAGTACAC---AGGCTACCCCTCTGCTGAGTTCAGAGACAGCTGAAGAG 4363
 Db 2487 TACTTACCTCATGTGTACAGAGAGAAATGGCATGATGATCAAGAAATCCCTCTGTGAAGAG 2546
 QY 4364 CTCATTCGGCATCTCTTCTTTGAGAAATTTCTCAATATACAGTGGCTCCATTTGGGT 4423
 Db 2547 AATACGACGTTCTGCTTTGTATTAACACATTCACACAGATTAAGGGAATGCTTGAAG 2606
 QY 4424 GTGATATTCCTGCGCTGATTTTGGGCTCTCACTAGAGATGCAATCCCTTATCTTC 4483
 Db 2607 AAATACATCCGCGCTGATCTGTGAGAGAGTGTACAGACTGGAATCATGCTTTCCTCA 2666
 QY 4484 GCTGTGCTGATGTTGTTTCCATATTTACTTTGGGTAATCTCGGCTGAGAACTGCATTA 4543
 Db 2667 AGTGTCTGCTTATTTTCTTACTTAAATGAGATGCTTCCACACCGCATTTCAAG 2726
 QY 4544 CCAATTTGCAAGAGAGATACAGTCACTGCTGATCTATCTTACTTACTTACATAATT 4603
 Db 2727 T---TCCGGAACAAGCCATTTTGAACATTTATGATGATCTTCCCTACCAACAACC 2783
 QY 4604 TGTTCCTGCTCTCCAGGAATATGGAATGAGTGAAGGCTTGTCCAGAGCGGTG 4663
 Db 2784 TCATTTGCTCTTTTCAAGAAATATGATGATTAATTAATTAATTAATTAATTAATTAATTA 2843
 QY 4664 CAGATCCCTGCTTACTAATGTTTGAAGCAAAAACCCGTGTCAGAGTACCTTAGAA 4723
 Db 2844 GTACAGTGAAGTTAAAGATATCTAGAAGGTGAAGAGATGCTATTAAGATTTCCAAAG 2903
 QY 4724 AAAGAAATAGTTGATAGAGTTCCTGATGACTATAGTCCCTCTGATCAAGCTTCTC 4783
 Db 2904 AATCTAACAAATTAATAACCTTCCAGAGATTACAGAGCCTCATTAATCAAGATCCA 2963
 QY 4784 ATTTCAGTGGCCGAGTGTGAGATGATGAGCGAAAGATCTCTCTGCTTTTCT 4843
 Db 2964 ATTTCTGCTGCCGAATTCAGTGTGATTAAGAGAGAGCCCACTCTGCTGTGT 3023
 QY 4844 GTGGGGCTATATCTATCTAGACATTTGCTGACAGAAATTTGGAACGGGGAAGAG 4903
 Db 3024 GCGGATCTCTGCTGCTGCCAGATTTCTGCTGCCAGCTGATCACTGGAAGGGAGAGATG 3083
 QY 4904 TTGGAAGTTGATTTTCAAGCACTTCACTGTGAGCCGAGTGTGCAATTTTCTTAAAA 4963
 Db 3084 TAGAGAGCTGCAACAGCTACACACTTCTGTGCGCTGTGAGATGGGCACTTCTCTGAGAG 3143
 QY 4964 TCAGAGAAATGCGGAGTGTCTGCTGTGAAGTTAAAGCAGAGAGCTGTGCTATCCAGCTC 5023
 Db 3144 TACGGGAATGTACAGTGTCTATTTTATGCTGCAAAAACCAAGGCTGTTTATTTCTCCTC 3203

QY 5024 CTTACTTGATGATATGAGAAACACACACCTCGCTGCAAGAGGGCAACCCCTTCATT 5083
 Db 3204 CTTCACCTTGATGACTATGGGAGACGACGAGGACTCAGAGGGGAATCTTTTACATT 3263
 QY 5084 TATCTCGTGTGAGCGTATCGGATCCATTTGCTGTGCAACACACATGATTTATAGAAG 5143
 Db 3264 TATGCAAAAGAGCGGATTTCAAGAAAGTTCAAGAGCTGTGCAACACACAGTGTCAAGAG 3323
 QY 5144 AGATTGCTAGAGCCAGAGACTATGATGATTTATTTGATTCGAATTCAGCA 5195
 Db 3324 AATTTGACATGACACAGAGAGCCAAATCAGACACGCTTGATGATGACTGGCA 3375
 RESULT 12
 AAA02411
 ID AAA02411 standard; cDNA; 733 BP.
 XX
 AC AAA02411:
 XX
 DT 19-MAY-2000 (first entry)
 XX
 XX Human colon cancer cell line polynucleotide sequence SEQ ID NO:2402.
 KW Human; colon cancer; tumour; diagnosis; gene expression product;
 KW Probe; detection; cancerous state; metastasis; identification;
 KW breast cancer; oestrogen receptor-positive breast cancer; therapy;
 KW oestrogen receptor-negative breast cancer; lung cancer; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO958675-A2.
 XX
 PD 18-NOV-1999.
 XX
 PF 13-MAY-1999; 99NO-US10602.
 XX
 PR 14-MAY-1998; 98US-0085426.
 PR 15-MAY-1998; 98US-0085537.
 PR 15-MAY-1998; 98US-0085696.
 PR 21-OCT-1998; 98US-0105234.
 PR 27-OCT-1998; 98US-0105877.
 XX
 PA (CHIR) CHIRON CORP.
 XX (HYSE-) HYSEQ INC.
 XX
 PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
 PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
 PI Lamson G, Drmanac R, Ctkvenjakov R, Dickson M, Drmanac S, Labat I;
 PI Leshkowitz D, Kita D, Garcia V, Jones LM, Stache-Crain B;
 XX
 DR MPI; 2000-126369/11.
 XX
 PT Polynucleotide library used to determine cancerous states of mammalian
 PT cells -
 PS
 PS Claim 1: Page 958; 1097pp; English.
 XX
 CC AAA00010 to AAA02716 represent polynucleotides isolated from cDNA
 CC libraries constructed from human colon cancer cell lines. The present
 CC invention also describes a method of detecting differentially expressed
 CC genes correlated with a cancerous state of a mammalian cell, comprising
 CC detecting at least one* differentially expressed gene product in a test
 CC sample derived from a cell suspected of being cancerous, where detection
 CC of the differentially expressed gene product is correlated with a
 CC cancerous state of the cell from which the test sample was derived.
 CC The polynucleotides sequences can be used in a method for detecting
 CC differentially expressed genes correlated with a cancerous state of a
 CC mammalian cell. The polynucleotides can also be used as probes for
 CC detecting and mapping related genes. They can be used in diagnosis and
 CC prognosis of diseases and disorders (e.g. identification of
 CC pre-metastatic or metastatic cancerous states, stages of cancer, or
 CC responsiveness of cancer to therapy). This is particularly for breast

```
CC cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
CC negative breast cancer, lung cancer, and colon cancer.
XX
SQ Sequence 733 BP; 218 A; 145 C; 152 G; 210 T; 8 other;

Query Match      12.5%; Score 652.8; DB 21; Length 733;
Best Local Similarity 96.2%; Pred. No. 5.9e-164;
Matches 688; Conservative 9; Mismatches 13; Indels 5; Gaps 3;

QY 664 YVCVNDHSHYDHGTCTATATACAGCTACAAAGAGCTCTTGACTGTGAGCTCGCAGAGGCC 723
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 19 CCAGGACACCGTCATATACAGCTACAAAGAGCTCTTGACTGTGAGCTCGCAGAGGCC 78
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 724 CAGTTGCATACCACTGCCATTGACAAAGAGGGTGTGCGGCTGTTAAAGCGGGAGCTTAT 783
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 79 CAGTTGCNTTCCACTGCCATTGACAAAGAGGGTGTGCGGCTGTTAAAGCGGGAGCTTAT 138

QY 784 GCTGCTGCCAGGACGAAGAGATATAAAGAGTCATTCAGAAAATGCTCTCAACAT 843
Db 139 GCTGCTGCCAGGACGAAGAGATATAAAGAGTCATTCAGAAAATGCTCTCAACAT 198

QY 844 CCACCTTCATGAGAAGTATTACACTCAGAGATTATGGCTCATCAGAAAATGCTTTCGGT 903
Db 199 CCACCTTCATGAGAAGTATTACACTCAGAGATTATGGCTCATCAGAAAATGCTTTCGGT 258

QY 904 CTTGGTTCCTGGATGAACAAAATATAGAGCTATTCAAAGTGAAGTATAGCAGATCTTTTGC 963
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 259 CTTGGTTCCTGGATGAACAAAATATAGAGCTATTCAAAGTGAAGTATAGCAGATCTTTTGC 317

QY 964 CAAGCATGCTTAGAGAGAACCTGACTCGGAGATCCCTGTCTCATACAGCAGGTTAATG 1023
Db 318 CAAGCATGCTTAGAGAGAACCTGACTCGGAGATCCCTGTCTCATACAGCAGGTTAATG 377

QY 1024 CTTTGGATGCAAGCTTTTAAAGGTGCCGTAAAGATCCCTTCATGAATTCATCTTCAGC 1083
Db 378 CTTTGGATGCAAGCTTTTAAAGGTGCCGTAAAGATCCCTTCATGAATTCATCTTCAGC 437

QY 1084 AGTTTTTTTATGGAGATGGAATACAAAAAATCTTTGCTATGGAATTTGGAAGTATTAT 1143
Db 438 AGTTTTTTTATGGAGATGGAATACAAAAAATCTTTGCTATGGAATTTGGAAGTATTAT 497

QY 1144 AAACAACCTGCAAGAAATATATACAGTGTGATCATCAGAGATATCTCTATACATGCA 1203
Db 498 AAACAACCTGCAAGAAATATATACAGTGTGATCATCAGAGATATCTCTATACATGCA 557

QY 1204 CTTTCACTTCAGATGTTTACTGTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1263
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 558 CTTTCACTTCAGATGTTTACTGTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 617

QY 1264 GTTATCTCTGCTACTGAACTCTGCTAGAACTTTTACCTGAGTACTTTGACAGG--A 1321
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 618 GTTATCTCTGCTACTGAACTCTGCTAGAACTTTTACCTGAGTACTTTGACAGG--A 677

QY 1322 ACAATAAATTCACCTCCAGGGTTATAGCC--AGGACAAATGGGAGAGATATAT 1374
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 678 CCAATAAATTCACCTCCAGGGTTATAGCC--AGGACAAATGGGAGAGATATAT 732

RESULT 13
ID AAA02327 standard; cdna; 756 BP.
XX
AC AAA02327;
XX
DT 19-MAY-2000 (first entry)
XX
DE Human colon cancer cell line polynucleotide sequence SEQ ID NO:2318.
XX
KW Human; colon cancer; tumour; diagnosis; gene expression product;
KW probe; detection; cancerous state; metastasis; identification;
KW breast cancer; oestrogen receptor-positive breast cancer; therapy;
KW oestrogen receptor-negative breast cancer; lung cancer; ss.
XX
```

```
OS Homo sapiens.
XX
PN WO9958675-A2.
XX
PD 18-NOV-1999.
XX
PF 13-MAY-1999; 99WO-US10602.
XX
PR 14-MAY-1998; 98US-0085426.
PR 15-MAY-1998; 98US-0085537.
PR 15-MAY-1998; 98US-0085696.
PR 21-OCT-1998; 98US-0105234.
PR 27-OCT-1998; 98US-0105877.
XX
PA (CHIR ) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
PI Lanson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
XX
WPI; 2000-126369/11.
XX
Polynucleotide library used to determine cancerous states of mammalian
cells -
XX
PS Claim 1; Page 916-917; 1097pp; English.
XX
CC AAA00010 to AAA02716 represent polynucleotides isolated from cDNA
libraries constructed from human colon cancer cell lines. The present
invention also describes a method of detecting differentially expressed
genes correlated with a cancerous state of a mammalian cell, comprising
detecting at least one differentially expressed gene product in a test
sample derived from a cell suspected of being cancerous, where detection
of the differentially expressed gene product is correlated with a
cancerous state of the cell from which the test sample was derived.
CC The polynucleotides sequences can be used in a method for detecting
differentially expressed genes correlated with a cancerous state of a
mammalian cell. The polynucleotides can also be used as probes for
detecting and mapping related genes. They can be used in diagnosis and
prognosis of diseases and disorders (e.g. identification of
pre-metastatic or metastatic cancerous states, stages of cancer, or
responsiveness of cancer to therapy). This is particularly for breast
cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
negative breast cancer, lung cancer, and colon cancer.
XX
SQ Sequence 756 BP; 219 A; 145 C; 154 G; 219 T; 19 other;
```

```
Query Match      12.2%; Score 634.6; DB 21; Length 756;
Best Local Similarity 94.2%; Pred. No. 4.4e-159;
Matches 669; Conservative 10; Mismatches 28; Indels 3; Gaps 3;

QY 661 KNRYYCVNDHSHYDHGTCTATATACAGCTTACAAAGAGCTTGTGACTGTGAGCTCGCAGAG 720
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 50 TCGGCACGAGACCGTCATATACAGCTTACAAAGAGCTTGTGACTGTGAGCTCGCAGAG 109

QY 721 GCCCAGTTGCATACCACCTGCCATTGACAAAGAGGGTCGTGCGGCTGTTAAAGCGGGAGCT 780
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 110 GCCCAGTTGCATACCACCTGCCATTGACAAAGAGGGTCGCGGCTGTTAAAGCGGGAGCT 169

QY 781 TATGCTGCTTGCAGGAGCAAGGAGATATAAAGAGTCATTTCAGAAAATGTCTCTCAA 840
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 170 TATGCTGCTTGCAGGAGCAAGGAGATATAAAGAGTCATTTCAGAAAATGTCTCTCAA 229

QY 841 CATCCACTTCATGTAGAAGTATTACACTCAGAGATTATGGCTCATCAGAAAATTTGCTTTG 900
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 230 CATCCACTTCATGTAGAAGTATTACACTCAGAGATTATGGCTCATCAGAAAATTTGCTTTG 289

QY 901 CGTCTTGTTCTCGATGATGAACAAAATTTAGCTATTTCAGTCACTTTAGGCAAGATCTTT 960
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 290 CGTCTNGGTTCTCGATGATGAACAAAATTTAGCTATTTCAGTCACTTTAGGCAATATCTTT 349
```

OY 961 TGGCAAGCATGCTTGGAGAGAACCTGACTGGAGAGATCCGTCTCATTAAGCAGTTA 1020
 DB 350 TGGCAAGCATGCTTGGAGAGAACCTGACTGGAGAGATCCGTCTCATTAAGCAGTTA 409
 OY 1021 ATGCTTTGGGATCAAAAGCTTTTAAAGGTGCCCGTAAAGTCCCTTCAATGATGATCTTC 1080
 DB 410 ATGCTTTGGGATCAAAAGCTTTTAAAGGTGCCCGTAAAGTCCCTTCAATGATGATCTTC 469
 OY 1081 AGCAGTTTTTTTATGAGATGGAATCAAAAACCTCTTCTGATGAAATTTGGAAGTAT 1140
 DB 470 AGCAGTTTTTTTATGAGATGGAATCAAAAACCTCTTCTGATGAAATTTGGAAGTAT 529
 OY 1141 TATAACAACCTGAGAGAAATATATCATGATGATGATGATGATGATGATGATGATGAT 1200
 DB 530 TATAACAACCTGAGAGAAATATATCATGATGATGATGATGATGATGATGATGATGAT 589
 OY 1201 GCACCTTCAGTTCAGATGTTTACTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
 DB 590 GCAC-CTTCAGTTCAGATGTTTACTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 647
 OY 1261 AATGTTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
 DB 648 AATGTTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 707
 OY 1321 AACATTAATTCACCTTCAGGCTTATGCGCAGAGACAATTTGGAGAGAT 1370
 DB 708 AACATTAATTCACCTTCAGGCTTAT-GCCNGACANATTTGGAGAGATT 756
 RESULT 14
 AAH07621
 ID AAH07621 standard; cDNA; 712 BP.
 XX
 AC AAH07621;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human cDNA clone (5'-primer) SEQ ID NO:4456.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 XX
 OS Homo sapiens.
 PN EP1074617-A2.
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000BP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI; 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 PS Claim 1; SEQ ID 4456; 2537pp + CD ROM; English.
 CC
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of

CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH05166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 712 BP; 229 A; 132 C; 156 G; 190 T; 5 other;
 Query Match 10.3%; Score 534.6; DB 22; Length 712;
 Best Local Similarity 89.1%; Pred. No. 2.5e-132;
 Matches 627; Conservative 10; Mismatches 18; Indels 49; Gaps 4;
 OY 500 AACCTGGAAGCAGAGTACTATTAAGAGAAATTCACGCTGCTGCTGCTGCTGCTGCTG 559
 DB 8 ACTGGGAAGAGCAGAGTACTATTAAGAGAAATTCACGCTGCTGCTGCTGCTGCTGCTG 67
 OY 560 TTGTCCAAAGCAGAGAAATATTTCTCTGATGATATTAATATGCTGATGATGATGAT 619
 DB 68 TTGTCCAAAGCAGAGAAATATTTCTCTGATGATATTAATATGCTGATGATGATGAT 127
 OY 620 GGAAG 663
 DB 128 GGAAG 187
 OY 664 -----YYCVNDHSHSYDHGTCTATATACAGCTTACAA 694
 DB 188 ATTGTGCTTTTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 247
 OY 695 GAGCTCTTGAGTGTGAGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 754
 DB 248 GAGCTCTTGAGTGTGAGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 307
 OY 755 GTGCTGGGCGTTAAAGCGGAGAGCTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 814
 DB 308 GTGCTGGGCGTTAAAGCGGAGAGCTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 367
 OY 815 AGAGTCATTCAGAAAGTCTCTCAACATCCACTTCATGATGATGATGATGATGATGATGAT 874
 DB 368 AGAGTCATTCAGAAAGTCTCTCAACATCCACTTCATGATGATGATGATGATGATGATGAT 427
 OY 875 TTATGCTCATCAGAAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 934
 DB 428 TTATGCTCATCAGAAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 487
 OY 935 ATTCAATGACTTTAGAGAGATCTTTTCCCAAGCATGCTTATGATGATGATGATGATGATG 994
 DB 488 ATTCAATGACTTTAGAGAGATCTTTTCCCAAGCATGCTTATGATGATGATGATGATGATG 547
 OY 995 AGAATCCCTGCTCATAGAGAGAGTAAATGCTTTGGAGATGCAAGAGCTTATTAAGAGG-CC 1053
 DB 548 AGAATCCCTGCTCATAGAGAGAGTAAATGCTTTGGAGATGCAAGAGCTTATTAAGAGG-CC 607
 OY 1054 CGTAGATCCCTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1112
 DB 608 CGTAGATCCCTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 667
 OY 1113 ACTCTTGTGATGAGATTT--GTGAGTATTTATTAACAGTCA 1154
 DB 668 ACTCTTGTGATGAGATTTTGTGAAGATTTATTAACAGTCA 711

RESULT 15

AAS70068
ID AAS70068 standard; cDNA; 777 BP.

XX AC AAS70068;

XX DT 13-FEB-2002 (first entry)

XX DE DNA encoding novel human diagnostic protein #5872.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR P-PSDB; ABG05881.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity

XX FS Claim 1; SEQ ID No 5872; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 777 BP; 194 A; 204 C; 199 G; 180 T; 0 other;

Query Match 8.7%; Score 450.8; DB 23; Length 777;
Best Local Similarity 99.6%; Fred. No. 6.9e-110;
Matches 452; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 3165 CACCCACAGTCAGTACTACTAGAAATGCTTGGTCTCTAAACGGGGTCCATCTGT 3224
Db CACCCACAGTCAGTACTACTAGAAATGCTTGGTCTCTAAACGGGGTCCATCTGT 383
Oy 3225 TACTGAAAGAGGTGCTGACGTCATCTTTGCCAAGAGAACAGAGGTGAAATAGA 3284
|||||

Db 384 TACTGAAAGAGGTGCTGACGTGCATCCTTTCCCAAGAACAGAGGCTGAAATAGA 443
Oy 3285 AAATAATGCCATGGTATTATTCGGCCCTGTGCCAGAAATCTACTGCTTAAACCCAGCACAG 3344
Db 444 AAATAATGCCATGGTATTATTCGGCCCTGTGCCAGAAATCTACTGCTTAAACCCAGCACAG 503
Oy 3345 GGGAAACCCATAGAACTCTCAGGAGAGCCCTAGACCCACTTTTTCATGGATCCAGACTT 3404
Db 504 GGGAAACCCATAGAACTCTCAGGAGAGCCCTAGACCCACTTTTTCATGGATCCAGACTT 563
Oy 3405 GGCATATGGAACCTTATACAGAGGCTGTGGTTCATGTAATGCAGCGAGTGTGCTGGCAGAA 3464
Db 564 GGCATATGGAACCTTATACAGAGGCTGTGGTTCATGTAATGCAGCGAGTGTGCTGGCAGAA 623
Oy 3465 GTATTTTGAAGCTGTACAGCTGAGCTCTCAGCAGCGCATTCATGTTGACCTTTTGGACTT 3524
Db 624 GTATTTTGAAGCTGTACAGCTGAGCTCTCAGCAGCGCATTCATGTTGACCTTTTGGACTT 683
Oy 3525 GGAAGTGGAGAAATATCTTTGGCCCTCTTTGCAAAATCTCTGTGCAATACTGTGTATCCCAT 3584
Db 684 GGAAGTGGAGAAATATCTTTGGCCCTCTTTGCAAAATCTCTGTGCAATACTGTGTATCCCAT 743
Oy 3585 TATTCCTTTGCAACCTCAAAAGATAAACAGTGAG 3618
Db 744 TATTCCTTTGCAACCTCAAAAGATAAACAGTAG 777

Search completed: September 27, 2003, 01:57:59
Job time : 840.105 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 27, 2003, 01:01:23 ; Search time 221.076 Seconds
(Without alignments)
10391.910 Million cell updates/sec

Title: US-09-724-126a-18
Perfect score: 5205
Sequence: 1 atggcgcgacgagagagctg.....taactcgcagcttactctga 5205

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapept 1.0

Searched: 569978 segs, 220691566 residues
Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/2/ina/5a_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5b_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6a_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6b_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4085.6	78.5	6395	2	US-08-982-956-1
2	4085.6	78.5	6395	3	US-09-228-317-1
3	996.2	19.1	1001	2	US-08-982-956-2
4	996.2	19.1	1001	3	US-09-228-317-2
5	57.4	1.1	7218	1	US-08-232-463-14
6	52.4	1.0	7218	1	US-08-232-463-14
7	44.4	0.9	7333	3	US-09-138-024-21
8	44.4	0.9	7333	4	US-09-404-066-21
9	44.4	0.9	7333	4	US-09-573-322-21
10	40.6	0.8	41708	4	US-09-470-512A-3
11	39.8	0.8	2861	1	US-08-299-953-1
12	39.8	0.8	2861	1	US-08-459-415-1
13	39.8	0.8	2861	4	US-09-066-687-1
14	39.8	0.8	2861	5	PCT-US95-11231-1
15	39.8	0.8	2861	1	US-08-299-953-2
16	39.8	0.8	3881	1	US-08-459-415-2
17	39.8	0.8	3881	4	US-09-066-687-2
18	39.8	0.8	3881	5	PCT-US95-11231-2
19	39.8	0.8	4376	1	US-08-119-125A-1
20	39.4	0.8	751	4	US-09-687-698-11
21	38.4	0.7	5703	4	US-09-280-590A-36
22	38	0.7	1785	4	US-09-601-198-156
23	37.8	0.7	1664976	4	US-08-916-421B-1
24	37.8	0.7	1664976	4	US-08-916-421B-1
25	37	0.7	1113	1	US-08-341-538A-1
26	37	0.7	1113	2	US-08-725-518-1
27	37	0.7	1618	3	US-08-889-108-1

c 28	37	0.7	1618	3	US-08-889-108-3	Sequence 3, Appl1
c 29	37	0.7	1618	3	US-08-120-601B-1	Sequence 1, Appl1
c 30	37	0.7	1618	3	US-08-120-601B-3	Sequence 3, Appl1
c 31	37	0.7	1618	5	PCT-US94-10358-1	Sequence 1, Appl1
c 32	37	0.7	1618	5	PCT-US94-10358-3	Sequence 3, Appl1
c 33	37	0.7	4875	1	US-08-460-739-1	Sequence 1, Appl1
c 34	36.8	0.7	1020	4	US-09-690-454-35	Sequence 35, Appl1
c 35	36.2	0.7	9636	1	US-08-323-170B-1	Sequence 1, Appl1
c 36	36.2	0.7	9636	4	US-08-954-441-1	Sequence 1, Appl1
c 37	36.2	0.7	19124	2	US-08-487-826B-13	Sequence 13, Appl1
c 38	36	0.7	99500	4	US-09-798-096-10	Sequence 10, Appl1
c 39	35.8	0.7	3349	4	US-09-336-447A-2	Sequence 2, Appl1
c 40	35.6	0.7	203001	4	US-09-734-674-3	Sequence 3, Appl1
c 41	35.6	0.7	2319	4	US-09-107-532A-2070	Sequence 2070, Ap
c 42	35.6	0.7	5252	4	US-09-340-620A-31	Sequence 51, Appl
c 43	35.4	0.7	1449	4	US-09-134-001C-1472	Sequence 1472, Ap
c 44	35.4	0.7	2135	3	US-08-430-286A-1	Sequence 1, Appl1
c 45	35.2	0.7	289	3	US-09-007-005-17	Sequence 17, Appl

ALIGNMENTS

```

RESULT 1
US-08-982-956-1
; Sequence 1, Application US/08982956
; Patent No. 5861312
; GENERAL INFORMATION:
; APPLICANT: Varshavsky, Alexander
; TITLE OF INVENTION: NUCLEIC ACID ENCODING MAMMALIAN UB1
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kevin M. Farrell, P.C.
; STREET: P.O. Box 999
; CITY: York Harbor
; STATE: ME
; COUNTRY: US
; ZIP: 03911
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/982,956
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Farrell, Kevin M.
REGISTRATION NUMBER: 35,505
REFERENCE/DOCKET NUMBER: CIT-2001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (207) 363-0558
TELEFAX: (207) 363-0528
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6395 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 115..5385
US-08-982-956-1
Query Match      78.5%; Score 4085.6; DB 2; Length 6395;
Best Local Similarity 86.6%; Pred. No. 0;
Matches 4567; Conservative 9; Mismatches 629; Indels 69; Gaps 4;
1 ATGGCGACGAGGAGCTGAGAGTCTGAGAGATGGAATGACCGCGGAGTTACCCAG 60
|||||

```


Db 4495 GATCTGTCTCCAGGCGCCGCTTGTCTGAGGCTGAAGAGGATAGTGAGGAGGCTCGCTGT 4554
Qy 4372 GCATCTCTCTTCTTGTGAGAAATTTCTCAATATACAACTGGCTCAATTTGGGTGTGATATT 4431
Db 4555 GCATCTGTCTTCTTGTGAGAGTGTGCGAGCACAGAGCGGCTCAGTGGGTGCGGTGT 4614
Qy 4432 CCTGGTGTATTTGTGGTCTCACTGAAGATGCATCAACCCCTTATCTTCGCTGTCT 4491
Db 4615 CCCGGTGTACCTGTGGCTCTCCCTGAGAACGGCATCACCCCTTACTCCGCTGTCT 4674
Qy 4492 GCATTTGTCTTCCACTATTTACTTGGGTAATCTCCGCTGAGAACTGCATACCAATCT 4551
Db 4675 GCATCTCTTCTTCCACTATTTACTTGGAGTAGCTCGGCTGAGAACTGTTCGCCAATCT 4734
Qy 4552 GCAGAGGAGAGTACAGTGCCTCTGTAGCTATCTATCTTTACCTACAAATTTGTTCCTG 4611
Db 4735 GCTGAAGGAGAATTCAGTGCCTCTGTAGCTATCTATCTTTACCCACAAATTTGTTCCTG 4794
Qy 4612 CTCCTCCAGGAATATTGGGATCTGTAAAGCCCTTGTCTCCAGAGCGGTGTGCAGATCT 4671
Db 4795 CTTTCCAGGAATATTGGATACCATTAAGGCCCTTACTACAGAGGTGTGTGGAGATCT 4854
Qy 4672 GCCTTACTAACTGTCTTGAAGCAAAACACCGTGGTCACTACCTACCCAGAAAGAAAT 4731
Db 4855 GCCTTACTAACTGTCTTGAAGCAAAAGTGTGTGGTCACTACCTACCCAGAAAGAAAT 4914
Qy 4732 AGTTGTAGAGCTCTCTGATGACTATAGCTGCCTCTGAATCAAGCTTCTCATTTACG 4791
Db 4915 AGTTGTAGAGCTCTCTGAGGACTTACAGCTGTCTTCTAAATCAGGCTTCTCACTTAG 4974
Qy 4792 TGCCACGGTGTGAGATGATGAGGAAAGCATCTGTCTCTCTGCTCTGCTTCTCTGGGGCT 4851
Db 4975 TGTCACGGTGTGAGATGATGAGGAAAGCATCTGTCTCTCTGCTCTGCTTCTCTGGGGCT 5034
Qy 4852 ATACTATCTCTCAAGACATTTGCTGCCAGAAATTTGAAAGGGAAGAGTTGAGCT 4911
Db 5035 ATCTGTCTCTCAAGACATCTGTTGCCAAGAAATAGTGAATGGGGAAGAGTTGGAGCG 5094
Qy 4912 TGCATTTTTCAGGACTTCACTGTGGAGCGGAGTCTGCATTTTCTTAAATCAGAGAA 4971
Db 5095 TGCGTTTTCATGCGCTTCAATGTGTGTGCTGAGTCTGCATTTTCTTAAATCAGAGAA 5154
Qy 4972 TGCCGAGTGTCTGTGTGAAGTAAAGCCAGAGCTGTGCTATPCCAGCTCTTACTTG 5031
Db 5155 TGCAGGTGTCTGTGTGAAGGAAAGCCAGAGCTGTGCTTACCCAGCCCTTACTTG 5214
Qy 5032 GATGAATATGAGAAACAGACCTTGGCTGAGAGGGCAACCCCTTCAATTTACTCTGT 5091
Db 5215 GATGAATATGAGAAACAGACCTTGGCTGAGAGGGCAACCCCTTCAATTTACTCTCG 5274
Qy 5092 GAGCGGTATCGAAGCTCCATTTGGTCTGGCAACACACTGCATTTATAGAGAGATTTCT 5151
Db 5275 GAGCGGTATCGAAGCTCCATTTGGTCTGGCAACAGCACTGCATTTATAGAGAGATTTCT 5334
Qy 5152 AGGAGCCAGAGACTTAATCAGATGTTATTTGATTTCAACTGGCAGTTACTGTGA 5205
Db 5335 CGGAGCCAGGAGACTAATCAGATGCTATTTGATTTAACTGGCAGTTACTGTGA 5388

RESULT 2

US-09-228-317-1

; Sequence 1, Application US/09228317

; Patent No. 6159732

; GENERAL INFORMATION:

; APPLICANT: Varshavsky, Alexander

; APPLICANT: Kwon, Yong Tae

; TITLE OF INVENTION: NUCLEIC ACID ENCODING MAMMALIAN UBRI

; NUMBER OF SEQUENCES: 2

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Kevin M. Farrell, P.C.

; STREET: P.O. Box 999

; CITY: York Harbor

; STATE: ME

; COUNTRY: US
; ZIP: 03911
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/228,317
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Farrell, Kevin M.
; REGISTRATION NUMBER: 35,505
; REFERENCE/DOCKET NUMBER: CIT-2001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (207) 363-0558
; TELEFAX: (207) 363-0528
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6395 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 115..5385
; US-09-228-317-1

Query Match 78.5%; Score 4085.6; DB 3; Length 6395;
Best Local Similarity 86.6%; Pred. No. 0;
Matches 4567; Conservative 9; Mismatches 629; Indels 69; Gaps 4;
Qy 1 ATGGCGGAGAGAGGCTGAGGCTACTGAGAGGATGAATATCAAGCGGAGTTACCCAG 60
Db 115 ATGGCGGAGAGAGAGGCTGAGGCTACTGAGAGGATGAATATCAAGCGGAGGCTCCCTG 174
Qy 61 ACCCTTCAGCGCTCTGGCATCTTGGTGGGATCAGCAAGTTGATTTTATCTGCTTCTTG 120
Db 175 GCCCGCAGCGCGCGCATCGTGGTGGGATCAGCAAGTTGATTTCTATCTGCTTCTTA 234
Qy 121 CATCATTTGGCAAAATTTGTTGCCAGAAATTTACTTTGCTGAAATGGACCCAGACTTGGAA 180
Db 235 CATCATTTGGCAAAATTTAGTGGCAGAAATTTATTTGCTGAGATGGACCCAGATTGGAA 294
Qy 181 AAGCAGGAGAAAGTGTACAAATGTCAATATTCACCTCCACTGGAATGGTACTTATTTGA 240
Db 295 AAGCAGGAGAGAGTGTACAGATGTCAATCTCACTCTTGGAGTGGTACTTATTTGA 354
Qy 241 GAAGATCCAGATATTTGCTTAGAGAAATTTGAACAGAGTGGAGCATTTTCAGCTTTGTGG 300
Db 355 GAGGATCCCGATATTTGCTTAGAGAAATTTAAACACAGTGGAGCGGTCCAGTTGTGGG 414
Qy 301 AGGGTTTTCAAAAGTGGAGAGCAACCTATTTCTGAGGGAATTTGCAATTTGATCCAAACA 360
Db 415 AAGGTTTTCAAAAGTGGAGAGCAACATATTTCTGTAGGGAATTTGCAATTTGATCCAAAC 474
Qy 361 TGTGTACTCTGTATGAGCTGTCTCCAGGAGAGTGTTCATATAAATCATCTGTACAGATG 420
Db 475 TGTGTCTCTGTATGAGCTGTCTCCAAAGTAGTGTTCATATAAACCCTTCGTTCAGAGATG 534
Qy 421 CATACTTCTACTGGAGGAGGTTCTGTGACTGTGGAGACACAGAGCATGGAACCTGGC 480
Db 535 CATACTTCTACTGGAGGAGGTTCTGTGACTGTGGAGACACAGAGCGTGGAAACCTGGC 594
Qy 481 CCTTTTGTGTAATCATGAACCTGGAAGAGCAGGTACTATAAAGAGAAATTCACGCTGT 540
Db 595 CCTTTTGTGTAATCATGAACCTGGAAGAGCAGGTACTATAAAGAGAGCTTACATTCG 654
Qy 541 CGTTGTAATGAAGAGGTAATTTGTCAGCCAGCAAAATATTTCTTCAGTGATATAATAT 600
Db 655 CCATTGTAATGAAGAGGTAATTTGTCAGCCAGGAGAAATATTTCCCTTCGGTGATATAATAT 714

Db	2875	GGCTTGTGGAAGAGACGACGCTTCAGAAAGCTCTGAAGAGNAAGTGCCTTTTGAC	2934	3955	AGTTTTGGAGTTCAGTCTTCGGTGAATATTTCAATAGTATCAAGGAAATGGTCAATCTC	4014
QY	2776	TTTTATCATAGAGCTTCAAGATTTGGGAATTCAGCCAT-----GAATATACAAATG	2826	3844	TTTGGCCACAACAATTTATAGAAATGGAATGCCACCTGATGAAAGGGATTCCTCGA	3903
Db	2935	TTTTACCATAAAGCTTCAAGATTTGGGAATTCAGCCATGAATGCTCAGAAATATACAAATG	2994	4015	TTTGGCCACAACAATTTATAGAAATTTGGCCCTGAAAGTGCCTCTGATGAACATGAGACCCACGA	4074
QY	2827	CTTTTGGAAAACCTCAAGGAATTTCCCAAGTTAGNAGCCAGAGGACATGATACGCTGG	2886	3904	GTCCCATCTGACCTGGAGCACCCTGCGCTTTTCACATATCCAGCAATTTGAAATCTATTG	3963
Db	2995	CTCTTGGAAAACCTCAAGGAATTTCCCAAGTTAGNAGCCAGAGGACATGATACGCTGG	3054	4075	GTGCCCATGATGACCTGGAGCAGCTGTGCTTCCACATCCAGCAATTCGAAACCTGTGTG	4134
QY	2887	ATACTTCAGATGTTTGCACACAGTGAAGCGATTAAAGAAAAATCTTTTAATTTAGTAC	2946	3964	GGAGATGAAGGAAAAACCTCTGTTTGGAGCACCTTCAAAATAGGAGCAGATATGCTCTGAAA	4023
Db	3055	ATACTCCAGATGTTTGCACACAGTGAAGCGATTAAAGAAAAATCTTTTAATTTAGTAC	3114	4135	GGAGATGAAGGAAAAACCTCTTATTGGAGCACCTTCAAAATAGAGCAGATACCGGTCTGAG	4194
QY	2947	ACCACATCAGATCGGAATCTATTAAAGAAATGATGAGATTACTCTATGATAAAGAAAA	3006	4024	GCATTAATGCAATTTGACAGTTGCACAGAGGATTAACCTGCTCCTCAGGTCTGTATACAGAA	4083
Db	3115	ACCACATTCAGGACTGGAGTCATTAAGAGTAGGAGATTACTCATGATAAAGAAAA	3174	4195	GGCCTAAATGCAGTTTGCAGTTGCACAGAGGGCTACCTGCGCTCAGGTCTGTATACACAA	4254
QY	3007	GAACGAAAAAGAAAGCTGAGCTAGGCTAGGCTACATCCCGAAGAGATGATGGCTCAGATG	3066	4084	CATCTGGTTCGCTCTCTATACAGTTGCTTCTTCAACATAAATCAAGAGATACACCATGC	4143
Db	3175	GAACGGAAGAAAGCTGAGCCGCTAGGCTTATCCCGAAGAGATCATGGCCAGATG	3234	4255	CATCTGGCTCGGCTCCTGTCAAGTTATCTTCTTAACCTGCAATCAGAAATACACGAGC	4314
QY	3067	TCTGCTTTACAGAAAACTTCATTGAAACTCATAACTCATGATGACAAATACATCAGAA	3126	4144	CTTCTGCTATAGATCTGTTTTCATGTTTGGTGGTGTGTTAGCATTTCCCATCTCCCTTG	4203
Db	3235	TCTGCTTTACAGAAAACTTCATTGAAACTCATAACTCATGATGACAAATACATCAGAA	3294	4315	CTTCTGCTGTGATCTCTTCCATGTTCTGTCGGCGCAGTCTTAGGTTCCCATCTCTTG	4374
QY	3127	ATGCTTGGGAAGAAAGATTCATTAAGAGGAGAGACACCCAGCAGTCACTGACATAC	3186	4204	TAATGGGATGACCCCTGTTGATCTGACGCTTCTTCAAGTTAGTTCTTCTTATACACCTT	4263
Db	3295	GTAACAGGAAGGAAGACTCCATTAATGAGGAAGAGACGCTCAGCAGTCACTGAGGCC	3354	4375	TATGGGATGACACCCCTGGATCTGACGCGTGCACCTTAGTTCTTTCATATAACCACTC	4434
QY	3187	TCTAGAATTCCTTGGTTCCTAAACGGGTCCTACTGTTACTGAAAGAGGTCGTGACG	3246	4264	TATCTCTTCCATTTGATACCATGGCACATGCTTTCAGATACCTACTTACAGTAGACA--	4321
Db	3355	TCTAGAATTCCTTGGGCCCCTAACGGGGCCCGCTGTTACCGAAAGAGGTCGTGACG	3414	4435	TATCTCTTCCATTCGTATCACCATGGCGCACATGCTTTCAGATACCTCTTACAACAGATA	4494
QY	3247	TGCATCTCTTCCAGAAAGACAGGAGGTGAAATAGAAAAATATGCCATGTTATATCG	3306	4322	-----CAGGCTAGCCCTGCTCAGGTTTCAAGAGACAGTGAAGAGGTCATCTCC	4371
Db	3415	TGCATCTCTGCAAGAGAACAGAGGTAAACTAGAAAAATATGCCATGTTATGTCAT	3474	4495	GATCTGCTCCAGGGCGCGCTGCTGAGGGTGAAGAGGATAGTGAAGAGGTCGCTGT	4554
QY	3307	GCCTGTGTCAGAAATCTACCTTAAACGAGCAGAGGGAAGAACCCATAGAACTCTCA	3366	4372	GCATCTCTTCTTTGACAGAAATTTCTCAATATACAAAGTGGTCCATTTGGGTGTGATAT	4431
Db	3475	GCATGTGTGAGAAATCCACGCCCTAACCCAGCAGACAGAGGAGCCCTGTGACCACTTA	3534	4555	GCATCTGCTCTTTTGTGGAAGTGTGCGACACACAGAGCGGCTCCTACCTGGGTGCGT	4614
QY	3367	GGAGAGCCCTAGACCCACTTTTCATGATCCAGACTTGGCATATGGAACATATACAGGA	3426	4432	CCTGCTGCTATTTGTGGGTCTCAGTGAAGATGGCATCACCCCTTATCTTCCCTGTGCT	4491
Db	3535	GGGAAACACTGGACCCCTCTTTTATGATCCAGACTTGGCACATGGAACATATACAGGA	3594	4615	CCCGCTGTGATCTGTTGGCTCTCCCTGAGAGAGGCGCATCACCCCTTACCTCCGCTGT	4674
QY	3427	AGCTGTGTCATGTAATGCACGACGTGTGCTGGCAGAAAGTATTTTGAAGCTGTACAGCTG	3486	4492	GCATTTTCTTCCACTTATTTACTTGGGTAACTCCGCTGAGGAACTGCATACCAATCT	4551
Db	3595	AGCTGTGTCATGTAATGCATGACGTGTGCTGGCAGAAAGTATTTTGAAGCTGTGAGCTG	3654	4675	GCATGCTTTTCCACTTATTTACTTGGAGTAGCTCCGCTGAAGAACTGTTTGCCAAATCT	4734
QY	3487	AGCTCTCAGCAGCGGATTCATGTTGACCTTTTGGACTTGGAAAGTGGAGATATCTTTGC	3546	4552	GCAGAGGAGAGTACAGTGCATCTGTAGTATCTATCTTACCTACAAATTTGTTCCTG	4611
Db	3655	AGCTCGCAGCAGCGCATTCAGTGTAGACCTGTTGACCTTGGAGCGGAGTACCTATGC	3714	4735	GCTGAAGGAGAAATTCAGTGCATCTGTAGTATCTATCTTTACCCACAAATTTGTCCCTG	4794
QY	3547	CCTCTTTCAGAAATCTCTGTGCAATCTGATCCCATTTATCTTTCACCTTCAACCTCA	3606	4612	CTCTTCCAGGAATATTGGGATACGTGAAGCCCTTGTCTCCAGAGGCGGTGTCAGATCT	4671
Db	3715	CCGCTCTGCAAGTCTCTCTGCAACACTGTCTATCCCATCTATCCCTTTTGCAGCCGAG	3774	4795	CTTTTCCAGGAATATTGGGATACCATGAGCCCTTACTACAGAGGTGGTGGAGATCT	4854
QY	3607	ATAACAGTGAATGCAATGCTCTTCTCAACTTTTGAACCTGGCAGGCTGGATACAG	3666	4672	GCCTTACTAACTGTTTGAAGCAAAAAACCCGCTGCTGAGTACCCCTAGAAAAAGAAAT	4731
Db	3775	ATCAACAGTGAATGCGAGGCTCTGCTCAACTTTTGAACCTTTGGCCCGGTGGATACAG	3834	4855	GCCTTACTCAAGTCTTTGAAGCAGAAAAAGTGTGTGCTGCTGAGTACCTTAGAAAAAGAA	4914
QY	3667	ACTGTTCGGCAGAAATACAGGTTATATAAGACATGCTTAAGAGGAAAAAC---CA	3723	4732	AGTTTGAAGAGCTTCTCTGATGACTATAGTGTGCTCCTGCTGCTGCTCTCTATTTTCAGG	4791
Db	3835	ACTGTCTTGCAGAAATATCGGTTATATAAAGCATGCTAAAGGAGAAAGCCCCAGCA	3894	4915	AGTTTGAAGAGCTTCTCTGAGGACTACAGTGTCTTCTAAATCAGGCTTCTCACTTTAGG	4974
QY	3724	ATTCTATTTTCTTAATCAAGGAATGGAGATCTACTTTGGAGTTCCATTCATCTG	3783	4792	TGCCACGCTCTGCAGATGATGAGCCAAAGCATCTCTCTGCTGCTCTCTCTCTCTCTCT	4851
Db	3895	GTTCCTGTCTTGTATTAATCAAGGAATGGGGATCAACTTTTGAAGTTTCAATTCATCTG	3954	4975	TGTCCACGCTCTGCAGATGATGAGCGAAAGCATCTCTGCTCTCTCTCTCTCTCTCTCT	5034
QY	3784	AGTTTGGCGTGTGCTTCCATTAATATTTCAATACCATCAAGGAATGGTTATTCTC	3843	4852	ATACTATGTTCTCAGAACATTTCTCTGCCAGGAATTTGTAAACGGGAAGAGGTTGGAGCT	4911
				5035	ATCTGTGTTCTCAGAACATCTGTTGCCAAGAAATAGTGAATGGGAAGAGGTTGGAGCG	5094

;; TITLE OF INVENTION: NUCLEIC ACID ENCODING MAMMALIAN UBRI
;; NUMBER OF SEQUENCES: 2
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Kevin M. Farrell, P.C.
;; STREET: P.O. Box 999
;; CITY: York Harbor
;; STATE: ME
;; COUNTRY: US
;; ZIP: 03911
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/228,317
;; FILING DATE:
;; CLASSIFICATION:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Farrell, Kevin M.
;; REGISTRATION NUMBER: 35,505
;; REFERENCE/DOCKET NUMBER: CIT-2001
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (207) 363-0558
;; TELEFAX: (207) 363-0528
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1001 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 1..999
;; US-09-228-317-2

Query Match 19.18; Score 996.2; DB 3; Length 1001;
Best Local Similarity 99.78; Pred. No. 3.3e-276;
Matches 998; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2059 ATGGATCCCAATAGTTCTTGTACTGTACTTCCAGAGGTATGAACCTGCGGAGGCTTTT 2118
DB 1 ATGGATCCCAACAGTTCTTGTACTGTACTTCCAGAGGTATGAACCTGCGGAGGCTTTT 60
QY 2119 AACAAAGACCATATCTACAAAGACCGAGGATTTGATTAAACAATAATACTACTAATAGAA 2178
DB 61 AACAAAGACCATATCTACAAAGACCGAGGATTTGATTAAACAATAATACTACTAATAGAA 120
QY 2179 GAATGCTTCAGGTCCTCATCTATATTGTTGGGTGAGCGTTATGTAACCTGGAGTGGGAAAT 2238
DB 121 GAATGCTTCAGGTCCTCATCTATATTGTTGGGTGAGCGTTATGTAACCTGGAGTGGGAAAT 180
QY 2239 GTGACCAAAAGAGGTGCAATGAGAGAAATCAATTCACCTGCTTCATGTAACCCCATG 2298
DB 181 GTGACCAAAAGAGGTGCAATGAGAGAAATCAATTCACCTGCTTCATGTAACCCCATG 240
QY 2299 CCACACAGTGCCATTGTCACAAATTTACCTGAGAAATGAAATGAACTGGCTTAGAG 2358
DB 241 CCACACAGTGCCATTGTCACAAATTTACCTGAGAAATGAAATGAACTGGCTTAGAG 300
QY 2359 AATGTCATAACAAAGTGGCCACATTTAAGAAACCCAGGTGTATCAGGCCATGGAGTTTAT 2418
DB 301 AATGTCATAACAAAGTGGCCACATTTAAGAAACCCAGGTGTATCAGGCCATGGAGTTTAT 360
QY 2419 GAACATAAAGATGAATCACTGAAAGACTTCAATATGTACTTTTATCATTACTCCAAACCC 2478
DB 361 GAACATAAAGATGAATCACTGAAAGACTTCAATATGTACTTTTATCATTACTCCAAACCC 420
QY 2479 CAGCATAGCAGGCTGAACATATATGCAAGAAAGAGGAGAAACAAAGAAAGATGAA 2538
DB 421 CAGCATAGCAGGCTGAACATATATGCAAGAAAGAGGAGGAGAAACAAAGAAAGATGAA 480

QY 2539 GCATTGCCGCCACCACCCTCTCTGAATTTCTCCCTGCTTTTCAGCAAAAGTGATTAAACCTT 2598
DB 481 GCATTGCCGCCACCACCCTCTCTGAATTTCTCCCTGCTTTTCAGCAAAAGTGATTAAACCTT 540
QY 2599 CTCAACTGTGATATCATGATGTACATTTCTCAGAGCCGATTTTGGAGGGCAATAGACACA 2658
DB 541 CTCAACTGTGATATCATGATGTACATTTCTCAGAGCCGATTTTGGAGGGCAATAGACACA 600
QY 2659 GATTCTAACTTGTGGACCGAAGGGATGCTTCCAAATGGCTTTTCATATTTCTGGCATTTGGGT 2718
DB 601 GATTCTAACTTGTGGACCGAAGGGATGCTTCCAAATGGCTTTTCATATTTCTGGCATTTGGGT 660
QY 2719 TTACTAGAAGAGACACACAGCTTCAAAAGCTCTGAAGAGAGTACATTTGACTTT 2778
DB 661 TTACTAGAAGAGACACACAGCTTCAAAAGCTCTGAAGAGAGTACATTTGACTTT 720
QY 2779 TATCATAGGCTTCAAGATTTGGGAAGTTTCAGCCATGAATATACAAATGCTTTTGGAAAAA 2838
DB 721 TATCATAGGCTTCAAGATTTGGGAAGTTTCAGCCATGAATATACAAATGCTTTTGGAAAAA 780
QY 2839 CTCAAAGGAATTTCCCAAGTTAGAGGCCAGAGGACATGATAACCTGGATCTTCAGATG 2898
DB 781 CTCAAAGGAATTTCCCAAGTTAGAGGCCAGAGGACATGATAACCTGGATCTTCAGATG 840
QY 2899 TTTGACACAGTGAACCGATTAGAGAAAAATCTTGTAAATTTAGTACCAACCATCAGGA 2958
DB 841 TTTGACACAGTGAACCGATTAGAGAAAAATCTTGTAAATTTAGTACCAACCATCAGGA 900
QY 2959 TCGGAATCTATTAAAGATGATGAGATTACTCATGATAAAGAAAAAGCAGACGAAAAAGA 3018
DB 901 TCGGAATCTATTAAAGATGATGAGATTACTCATGATAAAGAAAAAGCAGACGAAAAAGA 960
QY 3019 AAAGCTGAAGCTGTAGGTACATGCCAGAGAGATCATGGC 3059
DB 961 AAAGCTGAAGCTGTAGGTACATGCCAGAGAGATCATGGC 1001

RESULT 5
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:

```
;
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
; US-08-232-463-14

Query Match
Best Local Similarity 1.1%; Score 57.4; DB 1; Length 7218;
Matches 43; Conservative 219; Mismatches 195; Indels 0; Gaps 0;

QY 2122 AAGACCATATCTACAAAGACAGGAGTTGATTAAACAATATATACACTAATAGAGAA 2181
Db 1481 AATTACTATCTATGCAAGTAGTTAAGAGATAGAGAAATTTGGTACRRRRRRRRRR 1422

QY 2182 ATGCTTCAGTCTCATCTATATATGTGGGTGAGCGTTATGTACCTGGAGTGGGAAATGTG 2241
Db 1421 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1362

QY 2242 ACCAAGAGAGGTCAATAGAGAGAAATCACTTGTCTTGCATTGAACCATGCCA 2301
Db 1361 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1302

QY 2302 CACAGTGCCATGCCAAAATTTACCTGAGAAATGAAATATGAACATGCTTAGAAT 2361
Db 1301 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1242

QY 2362 GTCATAAACAAGTGGCCACATTTAAGAAACAGGTGTATCAGGCCATGGAGTTATGAA 2421
Db 1241 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1182

QY 2422 CTAAAGATGAATCACTGAAAGACTTCAATATGACTTTTATCATTACTCCAAACCCAG 2481
Db 1181 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1122

QY 2482 CATAGCAAGCTCAACATATGAGAGAAAGAGAGAAACAAGAAACAAGATGAAGCA 2541
Db 1121 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1062

QY 2542 TTGGCCGCCACCACTCTGTAATCTGCGCTGCTT 2578
Db 1061 AAGCTCCCTGACCTGACCAAGCTCGGAATTAAT 1025
```

RESULT 6

```
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
```

```
;
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
; US-08-232-463-14

Query Match
Best Local Similarity 1.0%; Score 52.2; DB 1; Length 7218;
Matches 48; Conservative 218; Mismatches 211; Indels 0; Gaps 0;

QY 3958 CTATTGGAGATGAAGAAAACCTCTGTTGGAGCACTTCAAAATAGGAGCAGATAATGGT 4017
Db 971 CCATCGAGTGGCGGTACTATAACTATTTCTCTGTTGCCATAGCTCACAGAATAA 1030

QY 4018 CTGAAGCATTAATCCAGTTTGCAGTTGCACAGAGATTACCTGCTCCTCAGTCCCTGATA 4077
Db 1031 TTCCGAGCTTGGCTGAGGAGCTGCGATYYYYYYYYYYYYYYYYYYYYYYYY 1090

QY 4078 CAGAAATCATGTTGCTGCTTCTATCAGTTGTTCTTCTTAACATAAAATCAGAAGATACA 4137
Db 1091 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1150

QY 4138 CCATGCTTCTGCTATAGATCTGTTTCATGTTTGGTGGGTGCTGTGTAGCATPCCCA 4197
Db 1151 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1210

QY 4198 TCCTTGATATGGGATGACCTGTTGATCTGACGCTTCTTCAAGTCTTCTTCTTATAAC 4257
Db 1211 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1270

QY 4258 CACCTTATCTCTTCCATTTGATCACCATGGCACACATGCTTCCAGATCTACTTACAGTA 4317
Db 1271 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1330

QY 4318 GACAGGCTTACCCCTTGTCTCAGGTTCAAGAGAGAGTGAAGAGGCTCATTCGCGATCT 4377
Db 1331 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1390

QY 4378 TCCTTCTTTCGAGAAATTTCTCAATATACAGTGGCTCCATCGGTGGTGTATTCCT 4434
Db 1391 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1447
```

RESULT 7

```
US-09-138-024-21
; Sequence 21, Application US/09138024A
; Patent No. 6004779
; GENERAL INFORMATION:
; APPLICANT: Bradley, John D.
; APPLICANT: Thompson, Craig M.
; APPLICANT: Moore, Jeffrey B.
; APPLICANT: Wobbe, C. Richard
; APPLICANT: Healy, Judith M.
```

```

; APPLICANT: Donnelly, Caroline E.
; TITLE OF INVENTION: REGULATED GENE EXPRESSION IN YEAST
; FILE REFERENCE: 0342/1D469US1
; CURRENT APPLICATION NUMBER: US/09/138, 024A
; CURRENT FILING DATE: 1998-08-21
; EARLIER APPLICATION NUMBER: 60/056,719
; EARLIER FILING DATE: 1997-08-22
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows version 3.0
; SEQ ID NO 21
;     LENGTH: 7333
;     TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
;     OTHER INFORMATION: Plasmid p2M197
US-09-138-024-21

```

	Query Match	0.98;	Score 44.4;	DB 3;	Length 7333;
	Best Local Similarity	57.0%;	Pred. No. 0.056;		
	Matches 81;	Conservative 0;	Mismatches 61;	Indels 0;	Gaps 0;
Qy	294	TTGTGGGAGGTTTTCAAAAGTGGAGAGACAACCTATTCTTCGAGGGAATGTGCAATTGA	353		
Db	3222	TTGTGGGAGGAATTCAAAATAGGGGAACCCCTGTATAGTGTGCATGAGTGTGGTTCGGA	3281		
Qy	354	TCCAACATGTGTACTCTGTATGGACTGCTTCCAGGACAGTGTTCATAAAAATCATCGTTA	413		
Db	3282	TGATACTTGTGCTTGTGATTTCATGTTTTATCCAAAAGATCATGTGAATCATCATGT	3341		
Qy	414	CAAGATGCATCTTCTACTGGA	435		
Db	3342	TTGTACCGATATATGTACTGAA	3363		

```

RESULT 8
US-09-404-066-21
; Sequence 21, Application US/09404066
; Patent No. 6365409
; GENERAL INFORMATION:
; APPLICANT: Bradley, John D.
; APPLICANT: Thompson, Craig M.
; APPLICANT: Moore, Jeffrey B.
; APPLICANT: Wobbe, C. Richard
; APPLICANT: Healy, Judith M.
; APPLICANT: Donnelly, Caroline E.
; TITLE OF INVENTION: REGULATED GENE
; FILE REFERENCE: 0342/ID4690S1
; CURRENT APPLICATION NUMBER: US/09/4
; CURRENT FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: US 09/138
; PRIOR FILING DATE: 1998-08-21
; PRIOR APPLICATION NUMBER: 60/056,71
; PRIOR FILING DATE: 1997-08-22
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Versi
; SEQ ID NO 21
; LENGTH: 7333
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Plasmid p2M197
US-09-404-066-21

```

	Query Match	0.9%	Score 44.4;	DB 4;	Length 7333;
	Best Local Similarity	57.0%;	Pred. No. 0.056;		
	Matches 81;	Conservative 0;	Mismatches 61;	Indels 0;	Gaps 0;
Qy	294	TTGTGGGAGGTTTTTCAAAAGTCGAGACACAACCTATTCTTCGAGGCGATTTGCCAATTGA	353		
Db	3222	TTGTGGGAGGAAAATTCAAATAGGGAAACCCTTGTTATAGGTGTCATCAGTGTGGTTCGGA	3281		
Qy	354	TCCAACATGTGTACTCTGTATGGACTGCTTCCAGGACAGTGTTCATAAAAAATCATCTGTTA	413		

```

Db      3282 TGATACGTGGTGCTTTGTTATTTCATGTGTTTTTAATCCAAAAGATCATGTGAATCATCATGT 3341
Qy      414 CAAGATGCATACTTCTACTGGA 435
          |   |   |   |   |   |   |
Db      3342 TTGTACCGATATATGTACTGAA 3363

RESULT 9
US-09-573-322-21
; Sequence 21, Application US/09573322
; Patent No. 6531289
; GENERAL INFORMATION:
; APPLICANT: Bradley, John D.
; APPLICANT: Thompson, Craig M.
; APPLICANT: Moore, Jeffrey B.
; APPLICANT: Wobbe, C. Richard
; APPLICANT: Bailey, David A.
; TITLE OF INVENTION: Regulated Gene Expression in Yeast and
; FILE OF INVENTION: Methods of Use
; FILE REFERENCE: 0342/1D469-US4
; CURRENT APPLICATION NUMBER: US/09/573,322
; CURRENT FILING DATE: 2000-05-18
; PRIOR APPLICATION NUMBER: 09/404,066
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 09/138,024
; PRIOR FILING DATE: 1998-08-21
; PRIOR APPLICATION NUMBER: 60/056,719
; PRIOR FILING DATE: 1997-08-22
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 21
; LENGTH: 7333
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: ZM197 plasmid
US-09-573-322-21

```

	Query Match	0.98;	Score 44.4;	DB 4;	Length 7333;
	Best local Similarity	57.08;	Pred. No. 0.056;		
	Matches 81;	Conservative 0;	Mismatches 61;	Indels 0;	Gaps 0;
Qy	294	TTGTGGGAGGGTTTTCCAAAAGTGGAGAGACAACTATTCTTGCAGGGATTGTGCAATTGA	353		
Db	3222	TTGTGGGAGGAANTTCAAAATAGGGGAACCTTGTATAGGTGTCATGAGTGTGGTTCGGA	3281		
Qy	354	TCCAACATGTGTACTCTGTATGGAGCTGCTTCCAGGACAGTGTTCATAAAAATCATCGTTA	413		
Db	3282	TGATACTGTGTGCTTTGTATTTCATTTGTTTTAATCCAAAAGATCATGTGAATCATCATGT	3341		
Qy	414	CAAGATGCATCTTCTACTGGA	435		
Db	3342	TTGTACCGATATATGTACTGAA	3363		

```

RESULT 10
US-09-470-512A-3/c
; Sequence 3, Application US/09470512A
; Patent No. 6376652
; GENERAL INFORMATION:
; APPLICANT: PhageTech, Inc.
; TITLE OF INVENTION: Compositions and methods involving an essential Staphylococcus
; TITLE OF INVENTION: gene and its encoded protein
; FILE REFERENCE: 21715/1000
; CURRENT APPLICATION NUMBER: US/09/470,512A
; CURRENT FILING DATE: 1999-12-12
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 41708
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-470-512A-3

```

Query Match	0.8%; Score 40.6; DB 4; Length 41708;	
Best Local Similarity	48.5%; Pred. No. 1.9;	
Matches 112; Conservative	0; Mismatches 119; Indels 0; Gaps 0;	
QY	2310 CATTTGCCAAATTTTACCTGAGAAATGAAATATGAAAGCTGGTTAGAGAAATGTCATAAA 2369	
DB	22756 CAGTCTCTCAAAATTTACTTTTAAACAAAATTTGATACAAAAGTAGTACATCAACAATTTTATTAA 22697	
QY	2370 CAAAGTGGCCACATTTTAAAGAAACCGGTGTATCAGGCCCATGGGTTTATGAACATAAAAGA 2429	
DB	22696 CAAATTTATCAACGTTATTTCTGCAATCAGGTAAAGCGTCACTTGGACATATGAAGAAGA 22637	
QY	2430 TGAATCACCTGAAGACTTCAATATGTTACTTTTATCTATCTTCAAAACCCAGCATAGCAA 2489	
DB	22636 AGCTATTAAATACGCCGTTTAAATTTTACAATTTATCCAAACGACACATATTAATAGCGT 22577	
QY	2490 GCGTGAACATATGCGAAGAAAGGAGAAACAAAGAAACAAAGATGAAGC 2540	
DB	22576 CACACTACCAAGAGAGTAGTAAAGCATAGATAGAAATATAGAAAAGAAAGC 22526	
RESULT 11		
US-08-299-953-1/c		
; Sequence 1, Application US/08299953		
; Patent No. 5646333		
; GENERAL INFORMATION:		
; APPLICANT: Dobres, Michael S. and Mandaci, Sevnuur		
; TITLE OF INVENTION: A Plant Promoter Useful for Directing the		
; TITLE OF INVENTION: Expression of Foreign Proteins to the Plant Epidermis		
; NUMBER OF SEQUENCES: 4		
; CORRESPONDENCE ADDRESS:		
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5646333ris		
; STREET: One Liberty Place 46th. Floor		
; CITY: Philadelphia		
; STATE: PA		
; ZIP: 19103		
; COMPUTER READABLE FORM:		
; MEDIUM TYPE: Floppy disk		
; COMPUTER: IBM PC compatible		
; OPERATING SYSTEM: PC-DOS/MS-DOS		
; SOFTWARE: PatentIn Release #1.0, Version #1.25		
; CURRENT APPLICATION DATA:		
; APPLICATION NUMBER: US/08/299,953		
; FILING DATE: Herewith		
; CLASSIFICATION: 435		
; ATTORNEY/AGENT INFORMATION:		
; NAME: Beardell, Lori Y.		
; REGISTRATION NUMBER: 34,293		
; REFERENCE/DOCKET NUMBER: NOVA-0003		
; TELEPHONE: 215-564-8960		
; TELEFAX: 215-568-3439		
; INFORMATION FOR SEQ ID NO: 1:		
; SEQUENCE CHARACTERISTICS:		
; LENGTH: 2861 base pairs		
; TYPE: nucleic acid		
; STRANDEDNESS: double		
; TOPOLOGY: linear		
; MOLECULE TYPE: DNA (genomic)		
; HYPOTHETICAL: NO		
; ANTI-SENSE: NO		
US-08-299-953-1		
Query Match	0.8%; Score 39.8; DB 1; Length 2861;	
Best Local Similarity	46.5%; Pred. No. 0.68;	
Matches 128; Conservative	0; Mismatches 147; Indels 0; Gaps 0;	
QY	2867 AGAAGACATGATAACCTGGATCTTTCAGATGTTTGACACAGTGAACCGATTAAAGAGAA 2926	
DB	1290 AAAAAGAACGGAGGAGTTAATTTTAGATTTTTTACACGTATTAATAAATTTATATCAA 1231	
QY	2927 AATCTTGTGTTTAAATTGTAGCAACCATCATCAGGATCGGAATCTATTAAAGATGATGAGATTA 2986	
DB	1230 AAAAATATCTTTTATATTAATAAATGGAAGATATAATTTATATTAATTAATAAAGAAAG 1171	

QY 2987 CTCATGATATAAGAAAAAGCAGAACGAAAAAGAGCTGAAGCTGCTAGGCTACATCGCC 3046
 Db 1170 ATAATAATAAATCTAGAGTTATATAAATAAACAATAATTAATCTCTTTTAAACATGTGA 1111
 QY 3047 AGAAGATCATGCTCAGATGCTGCTTACAGAAAAAAGCTTCAATGAAACTCATAAACTCA 3106
 Db 1110 AAATGATTATATTATGATATAATTTTTTTTCAAAACAACCAATAATAAAAAATGATAGG 1051
 QY 3107 TGTATGACAATACATCAAAATGCTGGGAAGAA 3141
 Db 1050 AGTATTATCATATGTCAGAAAATTAATATAAAGAA 1016

RESULT 13
 US-09-066-687-1/c
 ; Sequence 1, Application US/09066687
 ; Patent No. 6339185
 ; GENERAL INFORMATION:
 ; APPLICANT: Dobres, Michael S. and Mandaci, Sevnur
 ; TITLE OF INVENTION: A Plant Promoter Useful for Directing the
 ; TITLE OF INVENTION: Expression of Foreign Proteins to the Plant Epidermis
 ; NUMBER OF SEQUENCES: 7
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 6339185ris
 ; STREET: One Liberty Place 46th. Floor
 ; CITY: Philadelphia
 ; STATE: PA
 ; ZIP: 19103
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/066.687
 ; FILING DATE: Herewith
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Beardell, Lori Y.
 ; REGISTRATION NUMBER: 34,293
 ; REFERENCE/DOCKET NUMBER: NOVA-0003
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 215-564-8960
 ; TELEFAX: 215-568-3439
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2861 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; US-09-066-687-1

Query Match 0.8%; Score 39.8; DB 4; Length 2861;
 Best Local Similarity 46.5%; Pred. No. 0.68;
 Matches 128; Conservative 0; Mismatches 147; Indels 0; Gaps 0;
 QY 2867 AGAAGGACATNATACGCTGGATCTTACAGATCTTTGACACAGTGAAGCGATTAAAGAA 2926
 Db 1290 AAAAAGAACGGAGGAGTTAATATTTTACACGCTATTAATAAAAAATATATCA 1231
 QY 2927 AATCTGTTTAAATGTTAGCAACCATCAGGATCGGAATCTATTAGAGATGATGATTA 2986
 Db 1230 AAAATATCTTTATATATTAATAAATAAATGGAAGAGATAAATTTATATAATTAATAA 1171
 QY 2987 CTCATGATAAAGAAAAAGCAGAACGAAAAAGCTGAAGCTGCTAGGCTACATCGCC 3046
 Db 1170 ATAATAATAAATCTAGAGTTATATAAATAAACAATAATTAATCTCTTTTAAACATGTGA 1111
 QY 3047 AGAAGATCATGCTCAGATGCTGCTTACAGAAAAAAGCTTCAATGAAACTCATAAACTCA 3106
 Db 1170 ATAATAATAAATCTAGAGTTATATAAATAAACAATAATTAATCTCTTTTAAACATGTGA 1111
 QY 3047 AGAAGATCATGCTCAGATGCTGCTTACAGAAAAAAGCTTCAATGAAACTCATAAACTCA 3106

Db 1110 AAATGATTATATTATCATATAATTTTTTTTCAAAACAACCAATAATAAAAAATGATAGG 1051
 QY 3107 TGTATGACAATACATCAGAAATGCTGGGAAGAA 3141
 Db 1050 AGTATTATCATATGTCAGAAAATTAATATAAAGAA 1016
 RESULT 14
 PCT-US95-11231-1/c
 ; Sequence 1, Application PC/TUS9511231
 ; GENERAL INFORMATION:
 ; APPLICANT: Dobres, Michael S. and Mandaci, Sevnur
 ; TITLE OF INVENTION: A Plant Promoter Useful for Directing the Expression
 ; TITLE OF INVENTION: of Foreign Proteins to the Plant Epidermis
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & Norris
 ; STREET: One Liberty Place 46th. Floor
 ; CITY: Philadelphia
 ; STATE: PA
 ; ZIP: 19103
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US95/11231
 ; FILING DATE: Herewith
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/299,953
 ; FILING DATE: September 2, 1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Beardell, Lori Y.
 ; REGISTRATION NUMBER: 34,293
 ; REFERENCE/DOCKET NUMBER: NOVA-0016
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 215-564-8960
 ; TELEFAX: 215-568-3439
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2861 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; PCT-US95-11231-1

Query Match 0.8%; Score 39.8; DB 5; Length 2861;
 Best Local Similarity 46.5%; Pred. No. 0.68;
 Matches 128; Conservative 0; Mismatches 147; Indels 0; Gaps 0;
 QY 2867 AGAAGGACATNATACGCTGGATCTTACAGATCTTTGACACAGTGAAGCGATTAAAGAA 2926
 Db 1290 AAAAAGAACGGAGGAGTTAATATTTTACACGCTATTAATAAAAAATATATCA 1231
 QY 2927 AATCTGTTTAAATGTTAGCAACCATCAGGATCGGAATCTATTAGAGATGATGATTA 2986
 Db 1230 AAAATATCTTTATATATTAATAAATAAATGGAAGAGATAAATTTATATAATTAATAA 1171
 QY 2987 CTCATGATAAAGAAAAAGCAGAACGAAAAAGCTGAAGCTGCTAGGCTACATCGCC 3046
 Db 1170 ATAATAATAAATCTAGAGTTATATAAATAAACAATAATTAATCTCTTTTAAACATGTGA 1111
 QY 3047 AGAAGATCATGCTCAGATGCTGCTTACAGAAAAAAGCTTCAATGAAACTCATAAACTCA 3106
 Db 1110 AAATGATTATATTATGATATAATTTTTTTTCAAAACAACCAATAATAAAAAATGATAGG 1051
 QY 3107 TGTATGACAATACATCAGAAATGCTGGGAAGAA 3141

Search completed: September 27, 2003, 15:55:48
Job time : 231.076 secs

2012 AAATGTATGATAAAGATATCATCATGCTTCAGATTGGTGCATCTTTAATGGATCCCAATA 2071

Db 128 AGATGTTTGACAAGGATGTAGTAATGCTTCAGACAGGTGTCTCCATGATGGATCCAAATC 187
Qy 2072 AGTTCTGTTACTGGTACTCTCAGAGGTATGAACATGTCGCGAGCTTTTAAACAAGACCATAT 2131
Db 188 ATTTCTTGATGATCATGCTCAGCGCTTTGAACTTTATCAGATTTTCACTACTCCAGACT 247
Qy 2132 CTACAAA-----GACCAGGATTTGATTTAAACAATATATACAC 2170
Db 248 ATGGAAAAGATTTAGTTCTGAGATTACCCATAGAGATGTGTTTCAGCAGAACATACTC 307
Qy 2171 TAATAGAAAATGCTTCAGTCTCATCTATATTTGTTGGGTAGCGTTATGCTAGCTGGAG 2230
Db 308 TAATAGAAAATGCTATACCTCATTAATGCTTTGTTGGAGAGATTTAGTCTCGGAG 367
Qy 2231 TGGGAATGTGACCAAAAGAGAGTCAATGATGAGAAAATCATTTCACTTGTCTTTGCAATG 2290
Db 368 TTGACAGGTAAATGCTACAGATCAAAATCAAGCGAGAGATTATCCATCAGTTGAGTATCA 427
Qy 2291 AACCCATGCCACAGTGGCATTTGCCAAAATTTTACCTGAGAAATGAAAATATGAACATG 2350
Db 428 AGCCTATGGCTCATAGTGAATTTGTTAAAGTCTTTTACCTGAAGATGAGAACAAGGAGACTG 487
Qy 2351 GCTTAGAGAATGTCATAAAACAAGTGGCCACATTTAAAGAACCAAGGTGTATCAGGCCATG 2410
Db 488 GCATGGAGGTGTAATCGAAGCACTTGCCCATTTCAAGAAACCTGGATTACAGGACGAG 547
Qy 2411 GAGTTTATGAATAAAGATGAATCACTGAAAGACTTCAATPATGACTTTTATCATTAAT 2470
Db 548 GCATGTATGAATGAAACCAAGATGTGCCAAAGATTTCAACTTGTATTTCTCATCTTTT 607
Qy 2471 CCAAAACCCAGCATAGCAAGCTCAACATATGACAGAGAAGAAAGAGAAAACA 2530
Db 608 CAAGGCGAGAACAGTCCAAAGCAGAGAAGCGCAACGGAATTTGAAAAGACAAAATAGAG 667
Qy 2531 AAGATGAAGATTCGCGCCACACCACTCTCTGAAATTTCTGCTGCTTTTCAGCAAGTGA 2590
Db 668 AAGATACAGACTCCCACTCCGCTGTGCTCTCATCTGCGCTCTGTTTTCAGACCTTGG 727
Qy 2591 TTAACCTTCTCAACTGTGATATCATGATGATACATTTCTCAGACCGTATTTGAGGCGGCAA 2650
Db 728 TTAACATTTTGCAGTCAGATGTCATGTTGTGCATCATGGGAACAATTTCTGCAATGGCTG 787
Qy 2651 TAGACAGATTTCAACTTGTGGACCGAAGGATGCTCCAAATGGCTTTTCAATATCTGG 2710
Db 788 TGGAAACATAATGGATATGCTGCTGTCAGAGTCCATGCTGCAAGGGTGTATCAATTTAATG 847
Qy 2711 CATTTGGTGTACTAGAGAAGCAACAGCTTCAAAAAGCTCTCTGAAG---AAGAAGTAA 2767
Db 848 GCATGGCACTACAGAAGAAAACACATTTAGAGAATGTCACGGAAGAGCATGTAGTAA 907
Qy 2768 CATTTGACTTTTATCAAGGCTTCAAGATTTGGGAAGTTACGCCATGAAT-----A 2818
Db 908 CATTTACTCTCACTCAGAAGATATCAAAACCTGTGGAAGCGCAAAAATTTCTCCTAGCA 967
Qy 2819 TACAAATGCTTTTGGAAAACCTCAAGGAATTTCCCAAGTTAGAGAGCCAGAGACATGA 2878
Db 968 TACTAGCTATGCTGGAAACACTACAAAATGCTCCCTACCTAGAGTTCACAAAGACATGA 1027
Qy 2879 TAACGTGGATACTTCAGATGTTTCACACAGTGAAGCGGATTAAGAGAAAATCTTGTATA 2938
Db 1028 TTCGTGGATATTAAGACATTTTAAATGCTGTAAAAGATGAGGGAGA-----GTTTAC 1081
Qy 2939 TTGTAGCAACACATCAGGATCGGAATCTATTAAAGATGATGAGATTACTCATGATAAAG 2998
Db 1082 CTACCACTCCGTCGACAGACAGCAAGGAACCATTAATGGAAGAGATTTCAAGGGACAAG 1141
Qy 2999 AAAAGCAGAACGAAAAGAAAGCTGAAGCTGTAGCTACATCGCCAGAGATCATG 3058
Db 1142 ACAAGCTGAGAGGAAGAAAAGCAGAGATTTGCCAGACTCGCCAGAGAAAAGATCATGG 1201
Qy 3059 CTCAGATGCTGCCCTTACAGAAAACCTTATTGAACTCATTAACATCATGTATGACATA 3118
Db 1202 CTCAGATGCTGTAATGCAGCGGATTTTATTGATGAACAAAAGAACTCTTTTCAGCAGA 1261

Qy 3119 CATCAGAAATGCTCGGAAAAGAGATTCATTTATGGAGAAGAGACACCCACAGCAGTCA 3178
Db 1262 CATTAAGAACTGGATGCTCAACCTCTGCTGTGTTCTTGATCA-----TAGCCCTGTGCTT 1315
Qy 3179 GTGACTACTCTAGAATTTGCTTTGGTCTTAACGGGGTCCATCTGTTACTGAAAAGAGG 3238
Db 1316 CAGATATGACACTTACAGCACTTGGGCCCCGCAAACTCAGGTTCCTGTAACAAAGACAAT 1375
Qy 3239 TGCTCAGCTGCATCTTTGCCAAGAAGAACAGGAGTGAATAATAGAAAATATATGCCATGG 3298
Db 1376 TCGTTACATGATATATGTTCAAGGAGCAAGAGTTAAAGTGAAGACAGGGCAATGG 1435
Qy 3299 TATTATCGCCCTGTCTCCAGAAATCTACTGCTTTAACCCAGCAGGGGAAAACCCATAG 3358
Db 1436 TCTTGGCAGCATTTGTTAGAGATCAACTGTATATCAAAAACACAGAAGTAAATTTATTC 1495
Qy 3359 RACTCTCAGGAGAAGCCCTAGACCCACTTTTTCATGATCCAGACTTGGCATATGGAACTT 3418
Db 1496 AAGATCCAGAAAATAT---GATCCATTAATTCATGCCCTCTGCTTGTGGAACAC 1552
Qy 3419 ATACAGGAAGCTGTGCTCATGTAATGCACGAGTGTCTGGCAGAGATTTTGTGAAGCTG 3478
Db 1553 ACACCTAGTAGCTGTGGCAGATTTGCAATGCCCATTTGTTGGCAAGGTATTTGATCCG 1612
Qy 3479 TACAGCTGAGCTCTCAGCAGCGCATTTGCTTGACCTTTT-----TGACTTGG 3526
Db 1613 TTCAAGCTAAAGAACAGCAGGCAACAGAGATTTAGCTTTACATACGAGCATGATGATAG 1672
Qy 3527 AAGTGGAGAAATATCTTTCCTCTTTCGCAATCTCTGTCGAATCTGTCATCTGATCCCATTA 3586
Db 1673 AAAAGGAGATTTCTTTGCCCCCTTTGGAATGCTGTAGTAACTATGTTTCTCTCTG 1732
Qy 3587 TTCCCTTTCAGCTCAAAAGATAAAGCAGTGAAGTGCAGATGCTCTGCTCAACTTTTGA 3646
Db 1733 TGCCTT-----CTCCAAGAAATATTTTAAACACAGGTAAATTTTCAGACCAACCAA 1786
Qy 3647 CCTCGCAGGTGGATACAGCTGTCTGGCCAGAAATATCAGGTTATAATATAAGACATG 3706
Db 1787 ATCTGACTCAGTGGATTTAGAACAAATATCTCAGCAATAAAGCATTTACAGTTTCTTAGGA 1846
Qy 3707 CTAAAGGAGAAACCAATTCCTATTTCTTTAATCAAGGAATGGAGATTCTACTTTGG 3766
Db 1847 AAGAAAGAACTCTCTAATAATGCTCTACAAAAGAAATTCAGAAAATGGGATGAATATAC 1906
Qy 3767 AGTTCCATTCCTGAGTCTTGGCGTTGAGTCTTCGATTAATAATTTCAAAATAGCATCA 3826
Db 1907 AGCTCCCTGAAGGGTTCAGGCCCTGATTTTCGCTTAAGATCCCTTATTTCTGAGAGCATAA 1966
Qy 3827 AGGAAATGTTATCTCTTTGCCACAAATTTATAGAATTTGAGATTGAAGTGCACCTG 3886
Db 1967 AAGAAATGCTAACGACATTTGGAACCTGCTACCTACAAAGTGGGACTTAAAGGTTCAATCCA 2026
Qy 3887 ATGAAAGGATCCTCGAGTCCCATCTGACCTGGAGACCTGCGCTTTTCACTATCATCAGG 3946
Db 2027 ATGAGAGAGATCCTCGTGTTCCTCAATGTTGGGTAGCTGCGGCTACACCATCAAA 2086
Qy 3947 CAATTTGAAAATCTATTGGAGATGAAGAAAACCTCTGTTTGGAGCATTCAAAATAGGC 4006
Db 2087 GCATAGAAGAATTTGAGTGATGAAGATAACCATTTGTTGGTCTTTTACCTTGCAGAC 2146
Qy 4007 AGCATTAATGGTCTGAAAGCATTAATGCACTTTTGCAGTTGTCAGAGAGATTTACCTGCTC 4066
Db 2147 TGGATGACTGCTTAGGTCAATTGACGAGATTTGGCCAGCAGCACTGGACAGTGGCATCAG 2206
Qy 4067 AGGTCCTGATACAGAAAACATCTGGTCTGCTTCTATCATGTTGTTCTTCTTAACATAAAT 4126
Db 2207 TTTTCAGTGGTGAAGCAATTTTGTGTAACCTTTTTCATCACTGCTGCTGCTTAATGACAGCC 2266
Qy 4127 CAGAAATACACCATGCTCTGCTATAGATCTGTTTTCATGTTTGGTGGGTGCTGTGT 4186
Db 2267 ATGAGGAACCTTCCATGCATATTAGATATTGACATGTTTATTTATTTGGTGGGCTTGGTGC 2326

Db 668 AAGATACAGCACTCCCACTCCGGTGTGCTCCCATTTCTGCCCTCTGTTTGCAGGCTGG 727
QY 2591 TTAACCTTCTCAACTGTGATATCATGATGTACATTTCTCAGGACCCGATTTTGGAGCGGCAA 2650
Db 728 TTAACATTTTTCAGTCAGTCATGTTGTGCTCATGGGAACAATTCGCAATGGGTG 787
QY 2651 TAGACACAGATTTCACTTAACTTGTGGACCGAAGGATGCTCCAAATGGCTTTTCATATTCGG 2710
Db 788 TGGACATATGGATATCCCTGGTCAGAGTCCATGCTGCAAGGGTGTACATTTAATYG 847
QY 2711 CATTTGGGTTTACTAGAGAGAGCAACAGCTTTCAAAAAGCTCCTGAAG---AAGAAGTAA 2767
Db 848 GCATGGCACTACAAGAGAAACAAACATTTAGAGAAATGTACGGAAGAGCATGTAGTAA 907
QY 2768 CATTTGACTTTTATCATAGGCTTCAAGATTTGGGAAGTTTCAGCCATGAAT-----A 2818
Db 908 CATTTACCTTCACTCAGAGATATCAAAACCTGGTGAAGCGCAAAAATTTCTCTAGCA 967
QY 2819 TACAAATGCTTTTGGAAAACCTCAAGGAATTTCCCAAGTTAGAAAGCCAGAGGACATGA 2878
Db 968 TACTAGCTATGCTGGAACACTACAAATGCTCCCTACTAGAAATCCACAAGACATGA 1027
QY 2879 TAACTGTGATCTTACAGATTTTGGACACAGTGAAGCGGATTAAGAGAAAAATCTTTTAA 2938
Db 1028 TTCGCTGATATTGAAGACTTTTAATGCTGTAAAGATGAGGGAGA-----GTTCAC 1081
QY 2939 TTGTAGCAACCATCAGATCGGATCTATTAAGAAATGATGAGATTTACTCATGATARA 2998
Db 1082 CTACCAGTCCGTCGACAGACAGAAAGAACCTAATGGAAGAGATTTCAAGGGACAAG 1141
QY 2999 AAAAGCAGAGCAAAAAGAGCTGAAGCTGTAGCTTACATCCGACAGAGATCATGG 3058
Db 1142 ACAAGCTGAGGGAGAGAAAGCAGAGATGCGAGACTGCGCAGAGAAAGATCATGG 1201
QY 3059 CTCAGATCTGCTCCTTACAGAAAAACTTCATTGAAACTCATAACTCATGTATGACAATA 3118
Db 1202 CTCAGATCTGAAATGACGGGCATTTTATTGATGAAAAACAAGAACTCTTTTCAGCAGA 1261
QY 3119 CATCAGAAATGCTGGGAAAGAGATTTCCATTTATGGAGAGAGACGCCAGAGTCA 3178
Db 1262 CATTAGAACTGGATGCTTCAACCTCTGCTGTCTTTGATCA-----TAGCCCTGTGGCTT 1315
QY 3179 GTGACTACTCTAGAAATGCTTTGGTCTCTAAACGGGTCCATCTGTTACTGAAAAGGAG 3238
Db 1316 CAGATATCACACTTACACACTGGCCCGGCACAACTCAGGTTCTGACAAACACAT 1375
QY 3239 TGCTGACTGTGATCTTTGCCAAGAGAACAGGAGGTGAAATAGAAATAATATGCAATGG 3298
Db 1376 TCGTTACATGTATATTGTCTCAAGAGGAGCAAGAAAGTTAAAGTGGAAAGCAGGGCAATGG 1435
QY 3299 TATTATCGGCTGTGTCAGAAATCTACTGCTTTAACCCAGCACAGGGGAAACCCATAG 3358
Db 1436 TCTTGGCAGCATTTGTCAGAGATCACTGTATTAATCAAAAACAGAGTAAATTTATTC 1495
QY 3359 AACTCTCAGGAGAACCCCTAGACCCACATTTTCATGGATCCAGACTTGGCATATGGAACAT 3418
Db 1496 AAGATCCAGAAAAATAT---GATCATTTATTCATGCACCCCTGATCTGCTTGTGGAACAC 1552
QY 3419 ATACAGGAAGTGTGGTTCATGTAATGACGAGTGTGCTGGCAGAGTATTTTGAAGCTG 3478
Db 1553 ACACCTAGTAGTGTGGGCACATTTATGCATGCCCATTTGTGGCAAGGATTTTGTATTCG 1612
QY 3479 TACAGCTGAGCTCAGCAGCGCATTTGTTGACCTTTT-----TGACTGG 3526
Db 1613 TTCAAGCTTAAAGAACAGGAGGCAACAGAGATTTACGCTTACATACGAGCTATGATGAG 1672
QY 3527 AAAGTGGAGATATCTTTGCCCTCTTTGCAAACTCTGTGCAATCTGTGATGCCCATTA 3586
Db 1673 AAAAGGAGATTTCTTTTGGCCCTTTTGTGAATGCTTGAGTAATACTGTTATTCCTGCG 1732
QY 3587 TTCCTTTGCAACCTCAAAAGATAAAGAGTGAAGTGCAGATGCTGCTCAACTTTTGA 3646
Db 1733 TGCCTTT-----CTCCAAGAAATATTTTAAACAACAGGTTAAATTTTTCAGACCAACCAA 1786

QY 3647 CCTGSCGCGGTGATACAGACTGTTCTGGCCAGAAATATCAGGTTATAATATAAGACATG 3706
Db 1787 ATCTGACTCAGTGGATAGAACAAATATCTCAGCAAAATAAAGCATTTACAGTTTCTTAGGA 1846
QY 3707 CTAAGGAGAAACCCCAATTCCTATTCTTTTAAATCAAGGAATGGGAGATTTCTACTTTGG 3766
Db 1847 AAGAAGAAAGTACTCTCTAATAATGCTCTACAAAGAAATTCAGAAATGTGGATCAATTAC 1906
QY 3767 AGTTCCATTTCCATCTGAGTGTGGGCTTGGAGTCTTCGATTAATAATATCAAAATAGCATCA 3826
Db 1907 AGTCCCTGAAGGGTTTTCAGGCTGATTTTCGCTTAAGATCCCTTATTTCTGAGAGATAA 1966
QY 3827 AGGAATGCTTATCTCTTTGGCACAAACATTTATAGAATTTGATTTCAAAAGTGGCCCTG 3886
Db 1967 AAGAAATGTAAGAGCATTTTGGAACTTGCTACCTACAAGTGGAGTAAAGGTTTCATCCCA 2026
QY 3887 ATGAAAGGATCTCTCAGTCCCATGCTGACCTGAGCAGCCTGGCTTTTCACTATCCAGG 3946
Db 2027 ATGAAGAGATCTCTGCTGTTCCCATTAATGTGTTGGGTAGCTGCGGTACACCATCCAAA 2086
QY 3947 CAATTGAAAATCTAATTTGGGAGATGAAGAAACCTCTGTTTGGAGCACATTTCAAAATAGC 4006
Db 2087 GCATAGAAAAGAAATTTTGAAGTGAAGATAAACCAATTTGTTGGTCTTTACCTTGCAGAC 2146
QY 4007 AGCATAAATGCTGGAAGCATTAATGCAGTTTGCAGTTTGCACAGAGGATTTACCTGTCTC 4066
Db 2147 TGGATGACTGCTTTAGGTCATTTGACGAGATTTGCCGACGACACTTGGACAGTGGCATCAG 2206
QY 4067 AGGTCCTGATACAGAAACATCTGTTCTGCTCTATCATGTTTCTTCAACATAAAAT 4126
Db 2207 TTTCACTGCTGCAAGGACATTTTGTAAACTTTTTCATCCTGCTTAAAGACGCC 2266
QY 4127 CAGAAATACACCATGCTCTCTATAGATCTGTTTCATGTTTGGTGGGTGCTGTGT 4186
Db 2267 ATGAGAACTTCCATCATATTAAGATATGACATGTTTCAATTTATTTGGTGGGTGCTG 2326
QY 4187 TAGCATTTCCCATCTTGTATTTGGATGACCCCTGTTGATCTGACGCTTCTTCAGTTAGTT 4246
Db 2327 TTGCATTTCTGGTTGTCAGTG-----TCAGGATTTTTCAGGATCAGCC 2371
QY 4247 CTTCTTATAAACCACTTTTATCTCTCCATTTGATCACCATGGCACACATCTTCAATAC 4306
Db 2372 TTGGCACTGGAGACCTTTCATTTCCATCTGCTTACTATGGCACACATCATACAGATCT 2431
QY 4307 TACTTACAGTAGACAC---AGGCCTACCCCTGCTCAGGTTTCAAGAAAGACAGTGAAGG 4363
Db 2432 TACTTACCTCATGTACAGAAAGAGATGGCATCAAGAAAAATCCCCCTTTGTGAAGAG 2491
QY 4364 CTCATTTCCGATCTTCTTTTTCAGAAATTTCTCAATATACAAGTGGCTCCATTTGGT 4423
Db 2492 ATTCAGCAGTCTTGTGTTGATPAAACACTTCACCATGATATCGGGAAGTGGCTTGAAG 2551
QY 4424 GTGATATTCCTGGCTGGTATTTTGGTCTCTCACTGAAGAATGGCATCACCCCTTATCTTC 4483
Db 2552 AAATACCATCCGCTGGCATCTGTGGAGGAGTCTCAGAGCTGGAATCATCCCTTTCCTGA 2611
QY 4484 GCTGTGCTGATTTGTTTTCCTACTATTTTCTGGGTAACTCCGCTCAGGAACTGCATA 4543
Db 2612 AGTGTCTGCTTTATTTTTCATTAAATGGAGTTCTTCCCTCCCAAGGATTTCAAG 2671
QY 4544 CCAATTTCTCAGAAAGAGAGTACAGTGCCTCTGTAGCTATCTATCTTTACCTACAATTT 4603
Db 2672 T---TCCCTGGAACAAGCCATTTTGAACATTTATGTAGCTATCTTTCCTTACCACAAACC 2728
QY 4604 TGTTCCTGCTTCTCCAGGAATATTGGGATCTCTTAAGGCCCTTGTCTCAGAGGGGTGTG 4663
Db 2729 TCATTTGCTCTTTTCAAGAAATAGTGAATATGAATTTTCACTGATTTGAAAGTTGGTGC 2788
QY 4664 CAGATCCTGCTTACTTAACTGTTTGAAGCAAAAAAACCCGTGCTGAGTACCTTAGAA 4723
Db 2789 GTAACAGTGAAGTTAAAGATATCTAGAAGGTGAAAGAGATGCTATAGAATATATCCAGAG 2848

Query Match 8.4%; Score 438.4; DB 11; Length 505;
Best Local Similarity 99.8%; Pred. No. 2.7e-110;
Matches 439; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3834 GGTATTCTCTTCCGACAAATTTATAGAAATTTGGAATGGAAGTGCACCTGATGAAG 3893
DB 65 GGTATTCTCTTCCGACAAATTTATAGAAATTTGGAATGGAAGTGCACCTGATGAAG 124

QY 3894 GGATCCTCGAGTCCCGATGCTGACCTGGAGACCTGGCTTTCACTATCCAGGCAATGA 3953
DB 125 GGATCCTCGAGTCCCGATGCTGACCTGGAGACCTGGCTTTCACTATCCAGGCAATGA 184

QY 3954 AAATCTATTGGGAGATGAAGAAACCTCTGTTTGGAGCACTTCAAAATAGGAGCATAA 4013
DB 185 AAATCTATTGGGAGATGAAGAAACCTCTGTTTGGAGCACTTCAAAATAGGAGCATAA 244

QY 4014 TGCTCTGAAGCAATTAATGCAAGTTTGCAGTTGACAGAGGATTAACCTGTCCTCAGGTCCT 4073
DB 245 TGCTCTGAAGCAATTAATGCAAGTTTGCAGTTGACAGAGGATTAACCTGTCCTCAGGTCCT 304

QY 4074 GATACAGAAACATCTGTTGCTCTCTATCAGTTGTTTCCCTAACATAAAATCAGAAAGA 4133
DB 305 GATACAGAAACATCTGTTGCTCTCTATCAGTTGTTTCCCTAACATAAAATCAGAAAGA 364

QY 4134 TACACATGCTCTGCTCTATAGATCTGTTTCATGTTTGGTGGGCTGCTGTTAGCAT 4193
DB 365 TACACATGCTCTGCTCTATAGATCTGTTTCATGTTTGGTGGGCTGCTGTTAGCAT 424

QY 4194 CCATCCTCTGTTATGGGATGACCTGTTGATCTGCAGCTTCTTCAGTTAGTCTTCTCTA 4253
DB 425 CCATCCTCTGTTATGGGATGACCTGTTGATCTGCAGCTTCTTCAGTTAGTCTTCTCTA 484

QY 4254 TAACACCTTTATCTCTTCC 4273
DB 485 TAACACCTTTATCTCTTAC 504

RESULT 6

US-09-822-849A-532
; Sequence 532, Application US/09822849A
; Patent No. US20020045170A1
; GENERAL INFORMATION:
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fechtel, Kim
; APPLICANT: Agostino, Michael J.
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakar
; APPLICANT: Graham, James R.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6403
; CURRENT APPLICATION NUMBER: US/09/822,849A
; PRIOR FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: 60/195,582
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 598
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 532
; LENGTH: 972
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-849A-532

Query Match 4.8%; Score 249; DB 9; Length 972;
Best Local Similarity 56.8%; Pred. No. 1.1e-57;
Matches 478; Conservative 0; Mismatches 360; Indels 3; Gaps 1;

QY 4355 GTGAAGAGGCTCATTCGGCATCTCTTCTTTGAGAAATTTCTCAATATACAGTGGCT 4414
DB 15 GTGAAGAGGAATCAGCAGTCTTCTGCTTTGTATATAAACACTTCACCATATACGGGAATG 74

QY 4415 CCATTGGGTGTGATATTCCTGGCTGTGTAATTTGTGGTCTCAGTGAAGAAATGGCATCACCC 4474
DB 75 CCGTGAAGAAATACCATCCGGCTGGCATCTCTGGAGGAGTGTACAGAGCTGGAAATCATGC 134

QY 4475 CTTATCTTCGCTGTCTGCATTTGTTTCCACTATTTACTTGGGTAACCTCCGCCCTGAGG 4534
DB 135 CTTTCTGAAAGTGTCTGCTTTATTTTCACTTAAATGAGTTCCTTCCGCCACCCG 194

QY 4535 RACTGCATACCAATTTCTGCAGAGAGAGTACAGTGCATCTCTAGCTATCTATCTTTAC 4594
DB 195 ACATTCAAGT---TCTGGAAACAAGCCATTTTGAACATTTATGTAGCTATCTTTCCCTAC 251

QY 4595 CTACAAATTTGTTCTGCTCTCTCCAGGAATATTTGGGATCTACTGTAAAGCCCTTCTCTCAGA 4654
DB 252 CAACAACCTCATTTGCCCTTTTCAAGAAATAGTGAGATAATGAATTCCTGATTTGAAA 311

QY 4655 GCGGTGTGCAGATCCTGCTTTACTAAACTGTTTGAAGCAAAACACACCGTGGTCAGGT 4714
DB 312 GTTGTGCCGTAAACAGTGAAGTTAAAGATATCTAGAAGTGAAGAGATGCTATAGAT 371

QY 4715 ACCCTAGAAAAAAGAAATAGTTTGTAGAGCTTCTCTGATGACTATAGCTGCTCTCTGAATC 4774
DB 372 ATCCAAGAGAAATCTAACAAATTAATAAACCTTCCAGAGGATTAACAGCAGCTCATTAATC 431

QY 4775 AAGCTTCTCATTTTCAGGTGCCCGACCGTCTGCAGATCATCAGCAAAACATCTCTCTCT 4834
DB 432 AAGCATCCAATTTCTCGTGGCCGAAATCAGGTGGTGATTAAGAGCAGAGGCCCAACCTCTG 491

QY 4835 GCCTTTCTGTTGGGCTATACTATGTTCTCAGAAACATTTGCTGCCAGGAAATTTGTAACG 4894
DB 492 GCCTTGTGTGGGATCTCTGCTGCTCCAGAGTTACTGCTGCCAGACTGAACCTGGAAG 551

QY 4895 GGAAGAGGTTGGAGCTTGCAATTTTACGCACTTCTCAGTGTGGAGCCGGAGTCTGCATTT 4954
DB 552 GGGAGGATGTAGGAGCTGCAGCTGCACACCTACTCTCTGCTGCTGGAGTGGGCTCT 611

QY 4955 TCCTAAAAATCAGAAATGCCGAGTGGTCTGTTGAAAGTAAAGCCAGAGGCTGTGCTT 5014
DB 612 TCCTGAGAGTACGGGAATGTACAGGTGCTATTTTGTAGTGGCAAAACCAAGGCTGTTTT 671

QY 5015 ATCCAGCTCTTACTTGGATGAATATGGAGAAACACACCTGCTGGAAGGGGGAAC 5074
DB 672 ATTCCTCTCTTACCTTGTGATGACTATGGGAGACCCAGGAGCTCAGACGGGGAATC 731

QY 5075 CCCTTCATTTATCTGCTGAGCGGTATCGGAAGTCCATTTGGTCTGGCAACAACTGCA 5134
DB 732 CTTTACATTTATGCAAGAGCGGATTCAGAAGATTCAGAAGCTCTGSCACCAACACAGTG 791

QY 5135 TTATAGAAGAGATTTCTAGGAGCCCAAGAGACTAATCAGATGTTTATTTGGATTCACTGGC 5194
DB 792 TCACAGAGGAATTTGGACATGCACAGGAAGCAATCAGACACTGTTGTCATTGACTGGC 851

QY 5195 A 5195
DB 852 A 852

RESULT 7

US-10-071-766-7/c
; Sequence 7, Application US/10071766
; Publication No. US20020192678A1
; GENERAL INFORMATION:
; APPLICANT: Huel-Mei Chen
; TITLE OF INVENTION: GENES EXPRESSED IN SENESENCE
; FILE REFERENCE: PA-0043 US
; CURRENT APPLICATION NUMBER: US/10/071,766
; CURRENT FILING DATE: 2002-02-07
; NUMBER OF SEQ ID NOS: 144
; SOFTWARE: PERL Program
; SEQ ID NO 7
; LENGTH: 4573
; TYPE: DNA
; ORGANISM: Homo sapiens

FEATURE: misc.feature
NAME/KEY: Inocyte ID No. US20020192678A1 1330204.22
OTHER INFORMATION: Inocyte ID No. US20020192678A1 1330204.22
NAME/KEY: unsure
LOCATION: 3145-3168
OTHER INFORMATION: a, t, c, g, or other
US-10-071-766-7

Query Match 4.2%; Score 218.2; DB 13; Length 4573;
Best Local Similarity 57.5%; Pred. No. 1.1e-48;
Matches 440; Conservative 0; Mismatches 308; Indels 17; Gaps 2;
QY 4431 TCCTGGCTGGTATTTGGTCTCACTCAAGATGGCATACCCCTTATCTTCGCTGTC 4490
DB 4535 TCCGGCTGGCATCTGTGGAGAGTGTGAGAGTGTGAGATCATGCTTCCCTGAAGTGTTC 4476
QY 4491 TGCATTTGTTTCCACTATTTACTTTGGGTAACTCCGCTGAGGAACCTGCATACCAATTC 4550
DB 4475 TGCCTTATTTTTCATTACTTAATGGAGTTCCCTCCCAACCCGACATTCAGT---TCC 4419
QY 4551 TGCAGAGGAGTACAGTCACTCTGTAGTATCTATCTTTACCTACAAATTTGTTCT 4610
DB 4418 TGGAAACAAGCCATTTGAACATTTATGTAGTATCTTTCCCTACCAACAACCTCATTTG 4359
QY 4611 GCTCTTCCAGGAATATTTGGGATACGTAAAGCCCTTGTCTCCAGAGCGGTGTCAGATCC 4670
DB 4358 CTTTTCAGAAATAGTGAATAATGAATCACTGATTGAAGTTGGTCCGCTAACAG 4299
QY 4671 TGCCTTACTAACTCTTTGAAGCAAAACACCCGTGCTGAGTACCTCCCTAGAAAAAGAA 4730
DB 4298 TGAAGTTAAAGATATCTAGAGGTGAAGAGATGCTATAGATATCCAGAGATCTAA 4239
QY 4731 TAGTTTGTAGAGTCTCTGATGACTATAGTGCCTCCTGAATCAAGCTTCTCATTTTCAG 4790
DB 4238 CAATTAATAACCTTCCAGAGGATTACAGAGCCTCATTAATCAAGCATCCCAATTTCTC 4179
QY 4791 GTGCCCGAGTCTGCAGATGATGAGCGAAGCATCTGTCTCTGCTCTTCTGTGGGCG 4850
DB 4178 GT-----GGTGATAAGAGCAGAGAGCCCACTCTGTGCTTGTGTGGGATC 4133
QY 4851 TATATATGTTCTCAGAACATTTGCTGCCAGGAATTTGTAAGCGGAGAGGTTGGAGC 4910
DB 4132 TCTGTCTGTCTCCAGAGTACTGCTGCCAGACTGAACCTGGAAGGGAGGATGTAGGAGC 4073
QY 4911 TTGCACAGCTCACACCTACTCTGTGGCTCTGGAGTGGGATCTTCCCTGAGAGTACGGA 4013
DB 4072 CTGCACAGCTCACACCTACTCTGTGGCTCTGGAGTGGGATCTTCCCTGAGAGTACGGA 4013
QY 4971 ATGCCAGTGGTCTGGTTGAAGGTAAGCCAGAGGCTGTGCTATATCCAGTCTCTTACTT 5030
DB 4012 ATGTCAGGTGCTATTTTGTAGTGGCAAAACCAAGGCTGTTTATTTCTCTCTTACCT 3953
QY 5031 GGATCAATATGAGAAACAGACCTTGGCTGGAAGGGGCAACCCCTTCATTTATCTCG 5090
DB 3952 TGATGACTATGGGAGACCGACCGAGGACTCAGAGGGGAAATCCTTTTATTTATGCAA 3893
QY 5091 TGAGCGGTATCGGAAGTCCATTTGCTTGGCAACAACACTGCATATAGAGAGATTCG 5150
DB 3892 AGAGCGATTCAGAGATTCAGAGCTCTGGGACCAACACATGTCAGAGAGAAATTGG 3833
QY 5151 TAGGACCAAGAGACTAATCAGATGTTATTTGGATTCAACTGGCA 5195
DB 3832 ACATGCACAGGAAGCAATCAGACACTGTTGGCATTTGACTGGCA 3788

RESULT 8
US-10-027-632-260533
; Sequence 260533, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129

CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 260533
LENGTH: 578
TYPE: DNA
ORGANISM: Human
US-10-027-632-260533

Query Match 4.1%; Score 212.2; DB 13; Length 578;
Best Local Similarity 86.1%; Pred. No. 1.1e-47;
Matches 235; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
QY 3098 ATAAACTCATGTATGACATACATACATGCTGGGAAGAAGATTCATTCATTCATGAGG 3157
DB 226 ATCTACCAAAATATGAGCAAGTACCAAAACATGTCTCTCATTTCTGATCATTTCTT 285
QY 3158 AAGAGACACCCAGCAGTCACTGACTTCTAGAAATTTGTTGGTCTTAAACGGGTC 3217
DB 286 TTGGTAGACCCAGCAGTCACTGACTTCTAGAAATTTGTTGGTCTTAAACGGGTC 345
QY 3218 CATCTGTTACTGAAAAGAGGCTGCTGACGTGCATCTTCCCAAGAGAACAGAGGTGA 3277
DB 346 CATCTGTTACTGAAAAGAGGCTGCTGACGTGCATCTTCCCAAGAGAACAGAGGTGA 405
QY 3278 AATAGAAATAATGCTGTTATTCGGCTGTGTCAGAAATCTACTGCTTAAACCC 3337
DB 406 AATAGAAATAATGCTGTTATTCGGCTGTGTCAGAAATCTACTGCTTAAACCC 465
QY 3338 AGCAGAGGGGAAACCCATAGAACTCTCAGGAG 3370
DB 466 AGCAGAGGGGAAACCCATAGAACTCTCAGGAG 498

RESULT 9
US-09-918-995-27470
; Sequence 27470, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 27470
; LENGTH: 465
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(465)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-27470

	Query Match	1.2%.	Score 65;	DB 11;	Length 465;
	Best Local Similarity	49.6%;	Pred. No. 4.9e-07;		
	Matches 198;	Conservative	0;	Mismatches 195;	Indels 6;
					Gaps 1;
QY	3517	TTTGACATTGGAAGGTGGAGAAATATCTTTGGCCCTCTTTGGCAAACTCTGTGTGCAATACTGTG	3576		
Db	69	TATGATCTAGAAAACGGAGAATTCCTTTGGCTTCTTTGTGAATGCTTCGAGTAATACTGTT	128		
QY	3577	ATCCCCATTATTCCTTTTGGCAACCTCAAAAGATAAACAGTGAAGATGCAGATGCTCTTGCT	3636		
Db	129	ATTCC-----CTGTGCTGCTCTCTCCCAAGAAATATTTTAAACAACAGGTTAAATTTTCA	182		
QY	3637	CAACTTTTGACCCTGGCAGCGTGGATACAGACTCTCTCTGGCCCAAAATATCAGGTTATAAT	3696		
Db	183	GACCAACCAATCTGACTCAGTGGATTAGACATAATCTCAGCAAAATAAAGCATTCACG	242		
QY	3697	ATAAGACATGCTAAAGGAGAAAACCCCAATTCCTATTTTCTTTAATCAAGGAATGGGAGAT	3756		
Db	243	TTTCTTAGGAAAAGAAAGTACTCTCTTAATAATGCCCTCTACAAAGAATTCAGAAAATGTG	302		
QY	3757	TCCTACTTTGGAGTCCCAATTCATCCTCAGTTTGGCGTTGAGTCTTCGATTAAATATCA	3816		
Db	303	GATGAATACAGCTCCCTGAAGGTTTCAGGCCTGANTTTGCTCTAAGATCCCTTATCTCT	362		
QY	3817	AATAGCATCAAGGAAATGGTTATTCCTTTTGGCCACAACAATTTATAGAATTTGGATTGAAA	3876		
Db	363	GAGAGCATAAAGAAATGCTAACGACATTTTGGACTGCTACCTACAGGTGGGACTAAG	422		
QY	3877	GTGCCACCTGATGAAGAGGATTCCTCGAGTCCCCCATGTG	3915		
Db	423	GTTCATCCCAATGAAGAGGATCCCTCGTGTGCCCAATAAG	461		

RESULT 10
 US-09-908-975-8907
 ; Sequence 8907, Application US/09908975
 ; Publication No. US20030165843A1
 ; GENERAL INFORMATION:
 ; APPLICANT: SHOSHAN, Avi
 ; APPLICANT: WASSERMAN, Alon
 ; APPLICANT: MINTZ, Eli
 ; APPLICANT: MINTZ, Liat
 ; APPLICANT: FAIGLER, Simchon
 ; TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICING
 ; TITLE OF INVENTION: THAT POPULATE A TRANSCRIPTOME
 ; FILE REFERENCE: 36688-0005
 ; CURRENT APPLICATION NUMBER: US/09/908,975
 ; CURRENT FILING DATE: 2001-07-20
 ; PRIOR APPLICATION NUMBER: US 60/287,724
 ; PRIOR FILING DATE: 2001-05-02
 ; PRIOR APPLICATION NUMBER: US 60/221,607
 ; PRIOR FILING DATE: 2000-07-28
 ; NUMBER OF SEQ ID NOS: 32337
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 8907
 ; LENGTH: 60
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-908-975-8907

	Query Match	1.2%	Score 60	DB 12	Length 60
	Best Local Similarity	100.0%	Pred. No. 2.7e-06		
	Matches 60	Conservative 0	Mismatches 0	Indels 0	Gaps 0
Qy	2199	CTATTGTGGGTGAGCGTTATGTACTGTGGAGTGGGAAATGTGACCAAGAAAGAGGTCAC	2258		
Db	1	CTATTGTGGGTGAGCGTTATGTACTGTGGAGTGGGAAATGTGACCAAGAAAGAGGTCAC	60		

RESULT 11
US-10-027-632-8115
; sequence 8115, Application US/10027632

```

; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 8115
; LENGTH: 611
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(611)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-8115

```

```

Query Match      1.0%; Score 53.8; DB.13; Length 611;
Best Local Similarity 96.5%; Pred. NO. 0.00074;
Matches 55; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      3900  TCGAGTCCCATGCTGACCTGGAGCACCTTCGCGTTTCTACTATCCAGGCAATTGAAA 3956
Db      1    TCGAGTCCCATGCTGACCTGGAGCACCTTCGCGTTTCTACTATCCAGGCAATTGTTAA 57

```

```

RESULT 12
US-10-171-581-278
; Sequence 278, Application US/10171581
; Publication No. US20030104426A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: Linsley, Peter
; APPLICANT: Mao Mao
; TITLE OF INVENTION: Signature Genes in Chronic Myelogenous Leukemia
; FILE REFERENCE: 9301-157-999
; CURRENT APPLICATION NUMBER: US/10/171,581
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/298,914
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 366
; SEQ ID NO 278
; LENGTH: 535
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: Modified_base
; LOCATION: 1 ... 535
; OTHER INFORMATION: n = a,c,g, or t
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: N38966
; DATABASE ENTRY DATE: 2001-06-18
US-10-171-581-278

```

Query Match	0.8%;	Score 44.2;	DB 14;	Length 535;
Best Local Similarity	63.8%;	Pred. No. 0.31;		
Matches 67;	Conservative	0;	Mismatches 38;	Indels 0;
				Gaps 0;

QY 3367 GGAGAGCCCTAGACCCCTTTTCATGGATCCAGACTTGGCATATGGAACTTATACAGA 3426
| | | | |
Db 192 GCAGAAAATATGATCCATATTCAATGACCCCTGATCTGTCTTTGGAAACACACACTAGT 251
QY 3427 AGCTGTGCTCATGTAAATCAGCCAGCTGTGCTGGCAGAGTATTTT 3471
| | | | |
Db 252 AGCTGTGGCACCATTATGCATGCCCATTTGTCGCAAGGTAATGT 296

RESULT 13

US-10-311-455-251/c
; Sequence 251, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determination of Cytosine Methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 251
; LENGTH: 5198
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-251

Query Match 0.8%; Score 44.2; DB 12; Length 5198;
Best Local Similarity 49.8%; Pred. No. 1.6;
Matches 112; Conservative 0; Mismatches 113; Indels 0; Gaps 0;
QY 2313 TCCAAAATTTTACCTGAGATGAAATTAAGAACTGGCTTAGAGATGTTCATTAACAA 2372
| | | | |
Db 300 TACCATAAATCCCACTTAATAAATAACAAACGAACTCGCTCAAAAAAATAAAAAA 241
QY 2373 AGTGCCACATTTAAGAAACCCAGGTGTATCAGGCCATGGAGTTTATGAACATAAAGATGA 2432
| | | | |
Db 240 ATTAACATAAGCTAATATCAATACCTATATACCACTTAAATAAATAAATAAATAA 181
QY 2433 ATCACTGAAGACTTCAATATGTACTTTTATCATTACTCCAAACCCAGCATAGCAAGGC 2492
| | | | |
Db 180 AAATCACCTTAATCAAAACAAACAAATTAACATAAACCATAATACCATACCTACCTCC 121
QY 2493 TGAACATATGCAAGAAAGAGAGAAACAGAAACAAAGATGA 2537
| | | | |
Db 120 AACTTAATAACAAATAAATAAATAAATAAATAAATAAATAA 76

RESULT 14

US-09-895-913A-153
; Sequence 153, Application US/09895913A
; Patent No. US20020160456A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean Francois
; APPLICANT: Oomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20020160456A1 Helicobacter Polypeptides in the
; FILE OF INVENTION: Genome
; FILE REFERENCE: 06132/043002

; CURRENT APPLICATION NUMBER: US/09/895,913A
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 08/861,227
; PRIOR FILING DATE: 1997-06-24
; NUMBER OF SEQ ID NOS: 368
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 153
; LENGTH: 1024
; TYPE: DNA
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (115)....(921)
US-09-895-913A-153

Query Match 0.8%; Score 44; DB 10; Length 1024;
Best Local Similarity 50.0%; Pred. No. 0.55;
Matches 137; Conservative 0; Mismatches 135; Indels 2; Gaps 1;
QY 2924 AAAAACTCTTTTAAATTTAGTACCAACACATCAGATCGGAATCTATTAAAGATGATGAGA 2983
| | | | |
Db 275 AATATTCAGGATTAAACAATAGCGACAAGTTTATTATTAGCGCTTGTAGTGTGGTGATA 334
QY 2984 TTTACTCATGTAAGAAAG 3043
| | | | |
Db 335 TTGATAACACAGATAGAGATTAGAACAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAG 394
QY 3044 GCAGAGAGATCATGCTGCATCATGCTGCCTTTACAGAAAAAATCTTCTTAAATCATATAAC 3103
| | | | |
Db 395 CGNACAAGAGTGGGATAGAGACTGGACAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 454
QY 3104 --TCATGTATGACAATATCATCATGAGAAATGCTTGGGAAAGAGAGATTCATTTATGGAGGAGA 3161
| | | | |
Db 455 ATTAGTTAAAAAGCAGAACAAATTTGCCAAGAAATATCGCCCAATTTCTTTATGAAAA 514
QY 3162 GAGCAGCCAGCAGTCACTGAGTACTCTAGAAAT 3195
| | | | |
Db 515 AATTAGGAATTAAGGTGGCATTTGCTATAGAGT 548

RESULT 15

US-09-814-353-21640
; Sequence 21640, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21640
; LENGTH: 2372
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:


```
; NAME/KEY: misc_feature
; LOCATION: 2372
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-21640

Query Match      0.8%; Score 43.8; DB 12; Length 2372;
Best Local Similarity 53.0%; Pred. No. 1.1;
Matches 149; Conservative 0; Mismatches 117; Indels 15; Gaps 2;

QY 4819 AAGCATGCTGCTCTCCCTTTTCTGTGGGCTATACTATGTCTCAGAACATTTGCTGC 4878
Db      || ||||| ||||| ||||| || ||||| ||||| ||||| ||||| ||||| |||||
QY 990  AAAGATCCTGCTGTTGCCCTTGTGTGGTACTTTTCTATGCCCTGAAAGGACTTTGCTGC 1049
Db      || ||||| ||||| ||||| || ||||| ||||| ||||| ||||| |||||
QY 4879 CAGGAATTTGTGAACGGGGAAGAGTTGGAGCTTGCATTTTTCACGCACATTCACGTGGA 4938
Db      || || || || || || || || || || || || || || || || || || || ||
QY 1050 AAGCAACAAGTTACTGTGAA-----TGTGTACTGCACCTCAGAACTGTGGT 1097
Db      || || || || || || || || || || || || || || || || || || || ||
QY 4939 GCCGGAGTCTGCATTTTCCCTAAAAATCAGAGATGCCGAGTGGTCCCTGTTGAAGGTAAA 4998
Db      || || || || || || || || || || || || || || || || || || || ||
QY 1098 GCAGGACAGGTATTTTCCTTTTGTGATCAATGCATCGGTAAATATCATCATTCGAGGTCA- 1156
Db      || || || || || || || || || || || || || || || || || || || ||
QY 4999 GCCAGAGGCTGTGCTATCCAGCTCCCTTACTTGGATGAATATGGAGAAACAGACCCTGGC 5058
Db      || || || || || || || || || || || || || || || || || || || ||
QY 1157 --CCGCTTCTGCCCTCTGGGGTTCGGTGTATTGGATGCTCATGGAGAGAGACCGGGAT 1214
Db      || || || || || || || || || || || || || || || || || || || ||
QY 5059 CTGAAGAGGGGCAACCCCTTCATTTATCTCGTGAGCGGTA 5099
Db      || || || || || || || || || || || || || || || || || || || ||
QY 1215 CTTAGCGAGGCAACCTCTACATTTGTAAGGAAAGATA 1255
Db      || || || || || || || || || || || || || || || || || || || ||
```

Search completed: September 27, 2003, 20:27:15
Job time : 823.228 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 27, 2003, 00:42:14 ; Search time 6516.54 Seconds
(without alignments)
19412.893 Million cell updates/sec

Title: US-09-724-126a-18
Perfect score: 5205
Sequence: 1 atggcgagcagagagctgg.....tcaactggcagttactgtga 5205

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.*

1: em_estba.*

2: em_esthum.*

3: em_estin.*

4: em_estmu.*

5: em_estov.*

6: em_estpl.*

7: em_estro.*

8: em_estc.*

9: gb_est1.*

10: gb_est2.*

11: gb_est3.*

12: gb_est4.*

13: gb_est5.*

14: em_estfun.*

15: em_estom.*

16: em_estom.*

17: em_gss_hum.*

18: em_gss_inv.*

19: em_gss_pln.*

20: em_gss_vrt.*

21: em_gss_fun.*

22: em_gss_mam.*

23: em_gss_mus.*

24: em_gss_pro.*

25: em_gss_rod.*

26: em_gss_phg.*

27: em_gss_vrl.*

28: gb_gss1.*

29: gb_gss2.*

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1981.6	38.1	2561	11 AK089616	AK089616 Mus muscu
2	825	15.9	960	12 BM472160	BM472160 AGENCOURT
3	804.6	15.5	3641	11 BC044903	BC044903 Mus muscu
4	771.4	14.8	865	13 BQ233617	BQ233617 AGENCOURT

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

5	708.8	13.6	712	13 BX104087	BX104087
6	673	12.9	682	9 AI929033	AI929033 au64c10.y
7	664.8	12.8	668	13 BU618516	BU618516 UI-H-FH1-
8	656.6	12.6	756	10 BG534574	BG534574 602553425
9	628.6	12.1	3627	11 AK083320	AK083320 Mus muscu
10	620	11.9	797	12 BG862813	BG862813 602799074
11	596	11.5	774	13 BU955331	BU955331 AGENCOURT
12	585.2	11.2	695	14 BY756516	BY756516 BY756516
13	583.6	11.2	703	12 BM114217	BM114217 L0801D04-
14	577.4	11.1	3058	11 AK078173	AK078173 Mus muscu
15	576	11.1	729	13 BU703363	BU703363 UI-M-FD0-
16	571.6	11.0	782	12 BU086469	BU086469 602849734
17	560	10.8	560	14 CB159971	CB159971 K-EST0219
18	545.4	10.5	684	14 BY764809	BY764809 BY764809
19	521.2	10.0	528	13 BX119443	BX119443 BX119443
20	517.6	9.9	691	14 BY736138	BY736138 BY736138
21	512.2	9.8	606	10 BE589438	BE589438 195602 BA
22	511.4	9.8	752	14 CA313244	CA313244 UI-CF-FNO
23	507.8	9.8	633	10 BB622408	BB622408 BB622408
24	507.2	9.7	659	14 BY750548	BY750548 BY750548
25	497.4	9.6	719	13 BU623096	BU623096 UI-H-FL1-
26	494.4	9.5	496	13 BX281065	BX281065 BX281065
27	494	9.5	495	14 CB129142	CB129142 K-ESP0178
28	487.2	9.4	1105	13 BU751567	BU751567 CH5#001.A
29	485.4	9.3	641	9 AI361043	AI361043 qy03f11.x
30	482.4	9.3	506	2 HSM077106	Bx486919 Homo sapi
31	479.8	9.2	624	10 BG219270	BG219270 RST39023
32	477.8	9.2	481	9 AI192195	AI192195 qc92e08.x
33	473.6	9.1	609	29 CC200266	CC200266 RRC195 Ba
34	473.6	9.1	777	14 CB232501	CB232501 AGENCOURT
35	470	9.0	478	9 AA401319	AA401319 zu603Q04.r
36	453.8	8.7	737	13 BU461940	BU461940 603775907
37	451.2	8.7	465	12 BM508126	BM508126 1j39f04.x
38	447.4	8.6	522	10 BG382648	BG382648 298622 MA
39	446.6	8.6	830	14 CB588645	CB588645 AGENCOURT
40	446.4	8.6	520	10 BG382624	BG382624 298590 MA
41	446	8.6	640	14 CD368666	CD368666 UI-H-FT1-
42	445.2	8.6	710	14 CB519911	CB519911 UI-M-G10-
43	441.2	8.5	583	12 BI540789	BI540789 454188 MA
44	440	8.5	751	13 BU269526	BU269526 603504922
45	439.8	8.4	905	13 BU272014	BU272014 603509652

ALIGNMENTS

RESULT 1
AK089616
LOCUS
DEFINITION
AK089616 2561 bp mRNA linear HTC 05-DEC-2002
Mus musculus activated spleen cDNA, RIKEN full-length enriched
library, clone:F830005C07 product:ubiquitin protein ligase E3
component n-recognin 1, full insert sequence.
ACCESSION
AK089616
VERSION
AK089616.1 GI:26354612
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
Carninci, P. and Hayashizaki, Y.
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
PUBMED
11042159

REFERENCE
AUTHORS

3 Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kitsuunai,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujisake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

4 Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yananaka,I.,
Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
Kuehl,P., Lewis,S., Matsuo,Y., Nikaudo,I., Pesole,G.,
Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
Hofmann,M., Hume,D.A., Kamilya,M., Lee,N.H., Lyons,P.,
Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
Riing,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
Sato,K., Schonbach,C., Seva,T., Shibata,Y., Storch,K.F., Suzuki,H.,
Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilmimg,L.,
Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.,
and Hayashizaki,Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)
21085660
11217851

REFERENCE
AUTHORS

5 The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

6 (bases 1 to 2561)
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,W.,
Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.
Direct Submission
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration and Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)

TITLE
JOURNAL

cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Tissues were provided by Dr. John Todd (Dept. of Medical Genetics
Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome
Trust/WRC building Addenbrookes Hospital Cambridge) whose
assistance we gratefully acknowledge.

Please visit our web site for further details.

URL:http://genome.gsc.riken.go.jp/

URL:http://fantom.gsc.riken.go.jp/.

FEATURES
source

Location/Qualifiers
1. 2561
/organism="Mus musculus"
/mol_type="mRNA"
/strain="NOD"
/db_xref="FANTOM_DB:F830005C07"
/db_xref="taxon:10090"
/clone="F830005C07"
/tissue_type="activated spleen"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
14. >2560

CDS

/notes="unnamed protein product; putative
ubiquitin protein ligase E3 component n-recognin 1
(MGDI|GI:1277977, GI|NM_009461, evidence: BLASTN, 100%,
match=2560)"
/codon_start=1
/protein_id="BAC40933.1"
/db_xref="GI:26354613"
/translation="MADEMDGAERMDVSEPPPLAPQRPASWMDQVDFYATFLHLHLA
QLVPEIFYAEMDPDLKQESVQMSILTPLEWYLFGEEDPDICLEKLGKSGAFOLCGKV
FKSGETTYSCRDCAIDPTCVLCMDQFSSVHKHRYKMTSTGCGGCCDGDTEAWKTVG
PRCDVHEPRGACTKESLHCPLENEVIAOARRIEPSVKIVEMTWEEKEKELPELO
IREKNERYCVLFLNDEHSHYDVHYSIQRALDCELAELAHHTTAIDKEGRRAVKAGYV
ATCQEAEDIKSHSENVQHPHLHVSVMHQFALRGKALRGSMNMKNKIMSYSDSFROI
FQACVLEPGESENPCILSRMLMDARLYGARKILHELIFSSFFMEMEKYKILFAMEF
VKYQLOKREYISDDHERSISITALSQVMTVPTLARHLIEQONISVITVLELVP
EYLDRNNKFNFGYSODKLGKRVAVTCDLKYILISKPVITWRLRQAQFLGFRSLFKI
LTCMGMEEIRQVGQHIEVDPEAAIAIOMOLKNILLMFOEWCACDEDLVLVAYKE
CHKVNRCTNFMSSTKVVLQCLGSHLETSKYKVSDELVSILHPLSRLAGLHVRLSR
LGAISLHFEVFPQVEVLVEYLRCLVLAQVVAEMRRNGLSLISQVFFYQVVK
CREEMVYKDIIMIQIGASIMDPNKLFLQLRYELTDAFNKTIISTKDDIILKOYNLI
EEMLOVLIVIGERYVPGVGNVREEVINREITHLLCIEPMPHSATARNPENNET
GLENVINKVATFKKPGVSGHGVYKLEKDFNMVYFYHYSKTQHSKAEHM"

BASE COUNT 763 a 509 c 617 g 672 t

ORIGIN

Query Match 38.1%; Score 1981.6; DB 11; Length 2561;
Best Local Similarity 86.5%; Pred. No. 0;
Matches 2205; Conservative 9; Mismatches 289; Indels 45; Gaps 1;
QY 1 ATGGCGGAGAGAGAGGCTGGAGTACTGAGAGATGGAATCAGCGCGGAGTATACCCAG 60
||||| ||||| ||| || || ||||| ||||| ||||| ||||| ||||| |||||
Db 14 ATGGCGGAGAGAGAGTGGACGCGCGGAGAGGATGGACGCGCGGAGCTCCCGCT 73
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 ACCCTTCAGCGCTGCGCATCTTGTGGGATCAGCAAGTTCATTTTATCTGCTTTCTTG 120
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 74 GCGCCGCGCGCGCGGATCGTGTGGGATCAGCAAGTTCATTTCTATCTGCTTTCTTA 133
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 121 CATCATTTGGCACAATTTGGTGGCAGAAATTTACTTCTGCTGAAATGGACCCAGACTTGGAA 180
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 134 CATCATTTGGCACAATTTAGTGCCAGAAATTTATTTTCTGAGATGGACCCAGATTGGAA 193
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 181 AGCAGAGAGAAAGTGTACAAATGTCAATATTCCTCCACTGGAATGGTACTTATTTGGA 240
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 194 AAGCAAGAAGAGAGTGTACAGATGTCAATACTCACTCTCTTTGGAGTGTATTTATTTGGA 253
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 241 GAAGATCCAGATATTTGCTTAGAGAAATTCAGACAGTGGAGCATTTACGCTTTGTTGGG 300
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 254 GAGGATCCGGATATTTGCTTAGAGAAATTAACACACAGTGGAGCGTTCCAGTGTGTGGGG 313
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 301 AGGGTTTTCAAAAGTGGAGAGAACCTATTTCTTTCAGGAGTGTGTCAATTTGATCCCAACA 360
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 314 AAGTTTTTCAAAAGTGGAGAAACACATATTCCTGTAGGATGTGCAATTTGATCCCAAG 373
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 361 TGTGTACTCTGTATGACTGCTTCCAGGACAGTGTTCATATAAATCATCGTTACAAGATG 420
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 374 TGTGTGCTCTGTATGACTGCTTCCAAAGTAGTGTTCATAAAAACCATCGTTACAAGATG 433
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 421 CATACTTCTACTGGAGAGGGTTCTGTGACTGTGGAGACACAGCATGGAAACTTGGC 480
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 434 CATACTTCTACTGAGGGGCTTCTGTGACTGTGGAGACACAGAGCGTGAAACTGGC 493
Qy 481 CCTTTTGTGTAATCATGAACCTTGGAGACGAGTACTATATAAGAGAAATTCAGCGTGT 540
Db 494 CCTTTTGTGTAATCATGAACCTTGGAGACGAGTACTATATAAGAGAAATTCAGCTTGC 553
Qy 541 CCGTTGAATGAAGAGGTAATTTGTCACAGCCAGGAAATATTTCTTCCAGTCATATAAATAT 600
Db 554 CCATTGAATGAAGAGGTAATTTGTCACAGCCAGGAAATATTTCTTCCAGTCATATAAATAT 613
Qy 601 GTCGTAGAAATGACTATATGGGAAGAGGAAAGAACTTCCCTCGAAGTCCAGTAAGG 660
Db 614 ATTGTAGAAATGACTATATGGGAAGAGGAAAGAAATTTCCCTCGAAGTCCAGTAAGG 673
Qy 661 KNR-----YVCVNDHSHSDH 675
Db 674 GAGAAAAATGAACGATACTATTGTGCTCTTTCAACGATGAGCCATTCGTATGATCAT 733
Qy 676 GTCATATACAGCCTACAAAGAGCTTTGACTGTGAGCTCGCAGAGGCCAGTTCGATACC 735
Db 734 GTGATCTACAGTCTGCAGAGAGCTTAGATTGCGAGCTTGCAGAGGCACAGCTGCACAG 793
Qy 736 ACTGCCATTGACAAAGAGGCTGCGGGGCTTTAAAGCGGAGGCTTATGCTGCTTGCAG 795
Db 794 ACTGCCATCGACAAAGAGGCTGCGGGGCTGTCAAAGCAGGTGTATGCGCACCTGCCAG 853
Qy 796 GAAGCAAGGAATATAAGAGTCAATTCAGAAATGTCTCTCAACATCCACTTCATGTA 855
Db 854 GAAGCAAGGAATATAAGAGTCAATTCAGAAATGTCTCTCAACATCCACTTCATGTA 913
Qy 856 GAAGTATTACACTCAGAGATTATGGCTCATCAGAAATTTGCTTTCGCTTGGTCTCTGG 915
Db 914 GAGTGTGCTGCTCCGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 973
Qy 916 ATGAACAAATATGAGCTATTCAAGTGACTTTAGGAGAGTCTTTTGGCAAGCATGCCCT 975
Db 974 ATGAACAAATATGAGCTATTCAAGTGACTTTAGGAGAGTCTTTTGGCAAGCATGCCCT 1033
Qy 976 AGAGAAGACTGACTCGGAGAAATCCCTGCTCTATAAGCAGGTTAAATGCTTTGGAGTCA 1035
Db 1034 GTAGAAGAACTGGCTCTGAAATCCCTGCTTATAAGCAGACTAAATGCTTTGGAGTCA 1093
Qy 1036 AAGCTTTATAAGGTGCGCTTAAGATCCTTCATGAATGATCTTCAGCAGTTTTTTTATG 1095
Db 1094 AAAGCTTTATAAGGTGCGCTTAAGATCCTTCATGAATGATCTTCAGCAGTTTTTTTATG 1153
Qy 1096 GAGATGGAATACAAAAAATCTTTGCTATGGAATTTGTGAAGTATTTAAACAACATGCGAG 1155
Db 1154 GAGATGGAATACAAAAAATCTTTGCTATGGAATTTGTGAAGTATTTAAACAACATGCGAG 1213
Qy 1156 AAAGAATATATCAGTGATGATCATGACAGAGTATCTCTATAACTGCACTTTTCAGTTTCAG 1215
Db 1214 AAAGATFACATCAGCAGCAGCAGCAGAGAGCAGTCTCCATAACCGCCCTGTCGCTGCGAG 1273
Qy 1216 ATGTTTACTGTTCTCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1275
Db 1274 ATGCTCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1333
Qy 1276 ATTTACTGAACTCTGCTAGAGTTTACCTGAGTACTTGGACAGGAAATATAAATTCAC 1335
Db 1334 ATTTACTGAACTCTGCTAGAGTTTACCTGAGTACTTGGACAGGAAATATAAATTCAC 1393
Qy 1336 TTCCAGGTTTATAGCCAGGACAAATTTGGGAAGAGTATATGAGTATGAGTATGAGTATGAG 1395
Db 1394 TTCCAGGTTTATAGCCAGGACAAATTTGGGAAGAGTATATGAGTATGAGTATGAGTATGAG 1453
Qy 1396 TATATCTGATCAGCAACCCACAAATATGGACAGAAAGATTAAGAAATGCACTTCTCTGAA 1455
Db 1454 TATATCTGATCAGCAACCCACAAATATGGACAGAAAGATTAAGAAATGCACTTCTCTGAA 1513
Qy 1456 GGTTTTCGATCTTTTTTGAAGATTTCTACCTGTATGAGGGAATGGAAGAAATCCGAAAGA 1515
Db 1514 GGTTCGCTTCTTCTGAAGATTTCTACCTGTATGAGGGAATGGAAGAAATCAGAAGA 1573

RESULT 2
BM472160
LOCUS

BM472160

960 bp

mrna

linear

EST 05-FEB-2002

Qy 1516 ~CAGGTTGGCACAACACATTCGAAGTGGATCTCTGATTGGAGGCTGCCATTTGCTATACAGATG 1575
Db 1574 CAAGTTGGACAACACATTCGAAGTGGAGCTGACTGGAGGCTGCCATTCGCTATACAGATG 1633
Qy 1576 CAATTGAAGAATATTTTACTCATGTTTCCAAAGAGTGTGCTGTGCTGATGAAGAACTCTTA 1635
Db 1634 CAACATAAGAAATATTTTGTCTATGTTCCAAAGAGTGTGCTGTGCTGATGAAGATCTCTTA 1693
Qy 1636 CTTGTGGCTTATAAAGAATGTCAAAAGCTGTGATGAGGTGCGAGTACCAGTTTCATATCT 1695
Db 1694 CTGTGGCTTATAAAGAATGTCAAAAGCTGTAATCAGGTGCGAGTACAAATTTTCATGTCC 1753
Qy 1696 AGTACAACAGACAGTAGTACAATCGTCTGCACATAGTTTGGAAACAAAGTCCCTACAGATA 1755
Db 1754 AGTACAACAGACAGTAGTACAATCGTCTGCAGTCTAGTCTGGAACCAATCTCAAGATG 1813
Qy 1756 TCTGAGGATCTTTGTAAGCATACATCTGCCACTCTCTAGGACCCCTGCTGCTCTTCATGTA 1815
Db 1814 TCTGAGGACCTTGTAAAGCATACACCTGCCACTCTCTAGAACACTTCTGCTGCTCATGTA 1873
Qy 1816 CTTTAAAGCAGGCTGGTCTGTTTCAAGACTGCATGAATTTGCTGCTTTTGGAGACTTT 1875
Db 1874 CGTTTAAAGCAGACTAGTCTATTTCAAGACTGCATGAATTTGCTGCTTTTGACAGCTTT 1933
Qy 1876 CAAGTAGAGTACTAGTGGAAATATCCTTTACGTTGCTCTGCTGTTGGTGGTCCAGGTTGTT 1935
Db 1934 CAAGTAGAGTCTGCTGGTGGTACCCGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1993
Qy 1936 GCTGAGATCTGCGGAAGAAATGAGCTGCTCTATTATAGCAGCTGTTTATTTATACCAGAT 1995
Db 1994 GCTGAGATCTGCGGAAGAAATGAGCTGCTCTATTATAGCAGCTGTTTATTTATTCAGAT 2053
Qy 1996 GTTAAGTGCAGAGAAATGATGATAAAGACTATCATGCTTCAGATTCAGATTCGCTGCTCT 2055
Db 2054 GTTAATGCAGGAGAAATGTACGATAAAGATATCATGCTTCAGATTCGAGCATCT 2113
Qy 2056 TTAATGGATCCCAATAAGTCTTGTGTTACTTGGTACTTTCAGAGTATGAATTCGCGAGGCT 2115
Db 2114 ATAATGGATCCCAACAAGTCTTGTGTTACTTGGTACTTTCAGAGTATGAATTCAGTCT 2173
Qy 2116 TTTAACAGACCATATCTACAAAGACCAGGATTTGATTAACAATATAATACATAATA 2175
Db 2174 TTTAACAGACCATATCTACAAAGACCAGGATTTGATTAACAAGTATATAATACATAATA 2233
Qy 2176 GAAGAAATGCTTCAGTCTCATATATTTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2235
Db 2234 GAAGAAATGCTTCAGTCTCATATATTTGTTGGGAGAGCTTATGTACCTGGAGTGGGA 2293
Qy 2236 AATGTGACCAAGAGAGAGTCAACATGAGAGAAATCATTCACTTGCTTTGCTATGAAACC 2295
Db 2294 AATGTGACCAAGAGAGAGTCAACATGAGAGAAATCATTCACTTGCTTTGCTATGAAACC 2353
Qy 2296 ATGCCACAGTGGCTATGCCAAAATTTACCTGAGAAATGAATAATGAACCTGGCTTA 2355
Db 2354 ATGCCACAGTGGCTATGCCAAAATTTACCTGAGAAATGAATAATGAACCTGGCTTA 2413
Qy 2356 GAGAATGTCATAAACAAGTGCACATTTAAGAAACCAGGCTGTATCAGGCCATGAGATT 2415
Db 2414 GAGAATGTCATAAACAAGTGCACATTTAAGAAACCAGGCTGTATCAGGCCATGAGATT 2473
Qy 2416 TATGAACATAAAGATGAATCACTGAAAGACTTCAATATGTACTTTTTTATCTACTTCCAAA 2475
Db 2474 TATGAATGAAGATGAATCACTGAAAGACTTCAATATGTACTTTTTTATCTACTTCCAAA 2533
Qy 2476 ACCAGCATAGCAGGCTGAACATATGC 2503
Db 2534 ACACAGCATAGCAGGCTGAACATATGC 2561

DEFINITION AGENCOURT_6469668 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:5576383
5', mRNA sequence.
ACCESSION BM472160
VERSION BM472160.1 GI:18521202
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 960)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM12328 row: c column: 08
High quality sequence stop: 720.
FEATURES
source
Location/Qualifiers
1..960
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5576383"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_92"
/note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC library."
BASE COUNT 230 a 227 c 218 g 283 t 2 others
ORIGIN
Query Match 15.9%; Score 825; DB 12; Length 960;
Best Local Similarity 96.0%; Pred. No. 1e-158;
Matches 889; Conservative 0; Mismatches 31; Indels 6; Gaps 4;
QY 4184 TGTAGCATCCCACTCTGATGGGATGACCCCTGTTGATCGAGCCTTCTTCAGTTA 4243
DB 1 TGTAGCATCCCACTCTGATGGGATGACCCCTGTTGATCGAGCCTTCTTCAGTTA 60
QY 4244 GTTCTTCCATAAACACCTTTATCTCTCCATTGATCACCATGGCACACATGCTTCAGA 4303
DB 61 GTTCTTCCATAAACACCTTTATCTCTCCATTGATCACCATGGCACACATGCTTCAGA 120
QY 4304 TACTACTTACAGTACACACGCGCTACCCCTGCTCAGGTTCAAGAACACATGGAAGG 4363
DB 121 TACTACTTACAGTACACACGCGCTACCCCTGCTCAGGTTCAAGAACACATGGAAGG 180
QY 4364 CTCATTCGCGATCTCTTCTTTGCGAGAAATTCATATACAGTGGCTCCATTGGGT 4423
DB 181 CTCATTCGCGATCTCTTCTTTGCGAGAAATTCATATACAGTGGCTCCATTGGGT 240
QY 4424 GTGATATTCCGCGTGGTATTGTTGGGCTCTCACTGAAGAAATGGCATCACCCTTATCTTC 4483
DB 241 GTGATATTCCGCGTGGTATTGTTGGGCTCTCACTGAAGAAATGGCATCACCCTTATCTTC 300
QY 4484 GCTGTGCGCATTTGTTTCCACTATTTACTTGGGGTAACTCCGCGTGAAGAAATGCATA 4543
DB 301 GCTGTGCGCATTTGTTTCCACTATTTACTTGGGGTAACTCCGCGTGAAGAAATGCATA 360
QY 4544 CCAATTCTGCAGAGGAGATACAGTGCACCTCTGTAGCTATCTATCTTTACCTACAAATT 4603
DB 361 CCAATTCTGCAGAGGAGATACAGTGCACCTCTGTAGCTATCTATCTTTACCTACAAATT 420

QY 4604 TGTTCCTGCTCTTCCAGGAATATTGGATACTCTAAGGCCCTTGTCCAGAGCGGTGTG 4663
DB 421 TGTTCCTGCTCTTCCAGGAATATTGGATACTCTAAGGCCCTTGTCCAGAGCGGTGTG 480
QY 4664 CAGATCCTGCTCTTACTAACTGTTTGAAGCAAAAAACACCGTGGTCCAGTACCTAGAA 4723
DB 481 CAGATCCTGCTCTTACTAACTGTTTGAAGCAAAAAACACCGTGGTCCAGTACCTAGAA 540
QY 4724 AAGAATAAGTTTGTAGTAGAGTTCCTGTAGTACTATAGTCCCTCCCTGAATCAAGCTTCTC 4783
DB 541 AAGAATAAGTTTGTAGTAGAGTTCCTGTAGTACTATAGTCCCTCCCTGAATCAAGCTTCTC 600
QY 4784 ATTTCAGGTGCCACCGTCTGCAGATGATGAGCAAAAGCATCTGCTCTGCTTTTCT 4843
DB 601 ATTTCAGGTGCCACCGTCTGCAGATGATGAGCAAAAGCATCTGCTCTGCTTTTCT 660
QY 4844 GTGGGCTATATCTATCTCTCAGCAATTTGCTGCCAGGAAATTTGTAACGGGAACAGG 4903
DB 661 GTGGGCTATATCTATCTCTCAGCAATTTGCTGCCAGGAAATTTGTAACGGGAACAGG 720
QY 4904 TTGAGCTTGCATTTTTCACGCACTTCACTGTGGAGCCGAGTCTGCATTTTCTTAAAAA 4963
DB 721 TTGAGCTTGCATTTTTCACGCACTTCACTGTGGAGCCGAGTCTGCATTTTCTTAAAAA 780
QY 4964 TCAGAAATGCCAGTGG-TCCTGGTTGAAGTTAAAGCCAGAGGCTGT-CCCTATCCAGC 5021
DB 781 TCAGAAATGCCAGTGGTTCCTGGTTGAAGTTAAAGCCAGAGGCTGTCCCTATCCAGC 840
QY 5022 T-CTTACTTTGGATGAATATGGAGAA--ACACACCTGCTGCCCTGAAGAGGGGAACCC 5077
DB 841 TCCCTTACTTTGGATGATATGGAATAAACAGAACCTTGGCCCTGAAGAGGGGGCAACCC 900
QY 5078 TTCAATTATCTCGTGAGCGGTATCGG 5103
DB 901 CCTTTTCATTNATTCCTCTGGAGCGG 926
RESULT 3
BC044903
LOCUS
DEFINITION BC044903 3641 bp mRNA linear HTC 10-FEB-2003
Mus musculus, Similar to ubiquitin ligase E3 alpha-II, clone
IMAGE:3493115, mRNA.
ACCESSION BC044903
VERSION BC044903.1 GI:28279967
KEYWORDS HTC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Strausberg, R.
Direct Submission
Submitted (23-JAN-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: ang@bcm.tmc.edu
Gunnaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Lounseged, H.,
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
A.N., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAK Plate: 40 Row: b Column: 10
This clone has the following problem: frame shifted.

FEATURES

source
Location/Qualifiers
1. .3641
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C2BECH II"
/db_xref="taxon:10090"
/clone="IMAGE:3493115"
/tissue_type="Mammary tumor metastasized to lung. Tumor
arose spontaneously from a senescent normal mammary
(clonal) outgrowth infected with the virus MMTV."
/clone_lib="NCI CGAP_Lu29"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
BASE COUNT 989 a 899 c 906 g 847 t
ORIGIN

Query Match 15.5%; Score 804.6; DB 11; Length 3641;
Best Local Similarity 54.7%; Pred. No. 1.9e-154;
Matches 1928; Conservative 0; Mismatches 1509; Indels 88; Gaps 13;
QY 1720 TGTGGACATGATTTTGGAAACAAAGTCTCTACAGATATCTGAGGATCTTGTGAAGCATACAT 1779
DB 22 TGTGGACACTCGTGGGAACCATCAGATCTGTGTTCCCAAGAAAAGTTAGCATTCAC 81
QY 1780 CTGGCACTCTCAGGACCCCTGTGTCTTCATGATGATGTTAAGCAGGCTGGGTGCTGTT 1839
DB 82 CTCCCAATTTCTCGCTTGTGTCAGGTTTGCATGTTGTTAAGCAAAAAGTGAAGTGCA 141
QY 1840 TCAAGACTCCATGAATTTGTGCTTTTGGAGACTTTCAAGTAGAGGTACTAGTGGATAT 1899
DB 142 TATAAATTTCCAGAGCTCTTACCTCTAAGTGAATCAGGCCACCCATGTTGATAGACAT 201
QY 1900 CCTTTACGTTGCTGTGTTGGTGGCCAGGTTGTTGCTGAGATGTTGGCGAAGAAATGGA 1959
DB 202 CCTCTAGATGCTTGTCTATGCTCAAGTGCATGCTGGGATGTTGGAGAAGAAATGGC 261
QY 1960 CTGTCTCTATTAGCCAGGTGTTTATTACCAAGATGTTAAGTGCAGAGAAGAAATGTAT 2019
DB 262 TTCTCTCTAGTAAATCAGATCTATTACTACCATAATGTGAATGCAGGCGAGAGATGTT 321
QY 2020 GATAAAGATATCATGATCTCAGATTTGGTGCATCTTATAGATGCCAATAGTCTTG 2079
DB 322 GACAAGGACATATGATGTTTGCAGACAGGTGTTCCATGATGATGATCCAAACCACTTCT 381
QY 2080 TTACTGGTACTTCAGAGGTATGAACCTGCGAGGCTTTTAAACAAGACCATATATCAAAA 2138
DB 382 ATGATCATGCTCAGCGCTTTGAACCTCTATCAGCTCTTCAGACGCTGATATGGGAG 441
QY 2139 -----AGACCAGGATTTGATTAACAAATATATACACTTAATAGAA 2178
DB 442 AGATTCACTTCAGGTTTACCATAAGGAGCTGTTACGAGAACAACTCTGATCGAA 501
QY 2179 GAAATGCTTCAGGTCTCATCTATATATTTGGGTGAGCGTTATGATACCTGGAGTGGAAAT 2238
DB 502 GAGATGCTCTACCTCATCATCTATGCTTGTGGGAGAAAGATTCAACCTGSGGTTGGACAG 561
QY 2239 GTGACCAAGAGAGGTGCACATGAGAGAAATCATTCATCTGTTTGCATTAACCCATG 2298
DB 562 GTGGCTGCCACAGATGAATCAAGAGGGAGATTTATCCATCAGCTGAGCATCAAACTATG 621
QY 2299 CCACACAGTGCCTATGCCAAAATTTACCTGAGAATGAAAATTAATGAACTGCTTAGAG 2358
DB 622 GCTCAGAGGAGCTGTGAAGTCTCTGCTCTGAGATGAGAACANGAGACCGGCATGGAG 681
QY 2359 AATGTCATAAACAAGTGGCCACATTTAAGAAACACAGGTGTATCAGGCCATGGAGTTAT 2418
DB 682 ACCGTCATCGAGTCCGTAGCATTTCAAGAAACCTGGCCTCACAGGACGAGCATCTAT 741
QY 2419 GAACTAAAAGATGAATCACTGAAAGACTTCAATATGATCTTTTATCATTTACTCCAAACC 2478
DB 742 GAGCTGAAGCCAGAGTGTGCCAAAGAGTTCAACCTGTATTTTATCATTTTCTCCAGGCA 801

QY 2479 CAGCATACGAGGCTGAACATATCAGAGAAAGAGAGAAAACAAACAAACAAAGATGAA 2538
DB 802 GAGCAGTCCAAGGCGAGAGAAAGCTCAGCGGAAATTTGAAAAGAGAAAATAAGAAATACA 861
QY 2539 GCATTGGCGCCACCAACCACTCTCTGAATTTCTGCCCTGCTTTTTCAGCAAAAGTGAATTAACCTT 2598
DB 862 GCATCTCCTCTCGGCTTTGCCACCGTTCCTCCCTTTGTCGCGAGTCTGCTTACATC 921
QY 2599 CTCAACTGTGATATCATGATGATACATCTCAGGACCGTATTTTGGAGGGCAATAGACACA 2658
DB 922 TTGCAGTGTGACGTGATCTGATACATCATGGAACGATCCTGCAGTGGGCTGTAGAGCAT 981
QY 2659 GATTCTAACTTGTGGACCGCAAGGATGCTCAAAATGGCTTTTTCATATTTCTGCGCATGGGT 2718
DB 982 CACGGGTCTGCTGCTGTCAGAGTCCATGTACAGAGGTTGCTGCATTTGATCGGGATGGT 1041
QY 2719 TTACTAGAAGAGAAAGCAACAGCTTCAAAAAGCTCTCTGAAGAAGAAAGT---AACATTTGAC 2775
DB 1042 CTCAGGAAGAGAGACCACTTGGAGAAATGCCGTGGAAGGCGATGTCAGACCTTCACC 1101
QY 2776 TTTTATCATAGGCTTCAAGATTTGGAAAGTTTCAG-----CCATGAATATACAAATG 2826
DB 1102 TTCACACAGAGATTTCAAGCCTGGTGTGATGCACCACATAACTCCCCGAGCATCTTAGCT 1161
QY 2827 CTTTGGAAAACCTCAAGAGAAATTTCCCGAGTTTAGAGGCCAGGAAGACATGATACGTTG 2886
DB 1162 ATGCTGGAGACCTTTCAGAAAGCCCTCTCCCTGGAAAGCCACCAAGGACATGATCAGGTTG 1221
QY 2887 ATACTTTCAGATGTTTGCACAGTGAAGCGATTAAAGAGAAAATCTTGTGTTAATTTGAGCA 2946
DB 1222 TTGCTAAAGATGTTTAAATGCAATTAAGAAATTAAGAGA-----GTGTTTATCCAGCAGC 1275
QY 2947 ACCACATCAGATCGGAATCTATTAAAGATGATGAGATTACTCATGATAAGAAAAGCA 3006
DB 1276 CCCGTGCCGAGCGGAGGAAACCATATGAGAGGAGAGCTCAAGAGACAGGACAAAGCA 1335
QY 3007 GAACGAAAAGAAAGCTGAAGCTGTAGGTGATCATCGCCAGCAAGATCATGCTCAGATG 3066
DB 1336 GAGAGAAAAGAAAGCGGAGATGCCAGACTCGCCGCGGAGAAAGATCATGCCCCAGATG 1395
QY 3067 TCTGCTTACAGAAAACCTTCATTGAAACTCATAAATCATGTATGACAAATACATACAGAA 3126
DB 1396 TCTGAGATGTCAGCGGCACTTCATTGACGAAAACAAAGAGCTCTCCAGCAGACCCCTAGAG 1455
QY 3127 ATGCTGGGAAAGAGATTTCCATTATGAGGAAGAGAGACACCCAGCAGCTCAGTGACTAC 3186
DB 1456 CTGGACACCTCTGCTCTGCCACTCT-----TGACAGCAGCCCTCCCGTTTCAGACGCA 1509
QY 3187 TCTAGAATTTGGTCTTAAAGGGGTCCATCTGTTTACTGAAAAGGAGGTGCTGACG 3246
DB 1510 GCTCTTACAGCACTGGGCCAGCAGACAGAGTCCCTGAAACCCAGAGACAGTTTGTCAACC 1569
QY 3247 TGCATCTTTGCCAAGAGAAACAGAGGTGAAAATAGAAAATAATATGCCATGTTATATCG 3306
DB 1570 TGTATATATGTCGAAGAGGAGCAAGAGGTGACTGTGGGAAGCAGGCGGCTGCTTTGGCA 1629
QY 3307 GCCTGTCTCCAGAAATCTACTGCCTTAACCCAGCAGACAGGGGAAAACCCATAGAACTCTCA 3366
DB 1630 GCATTTGTCAGAGGTCAACCGTCTGTCAAAAGACAGAGCAAGAAAACCATCCGGACCCA 1689
QY 3367 GGAGAAGCCCTAGACCCACTTTTTCATGATCCAGACTTGGCATATGGAATTTATACAGGA 3426
DB 1690 GAAAAA---TATGATCCATTAATTCATGACCCCGATCTGTCTTGTGGGACACACACTGGC 1746
QY 3427 AGCTGTGCTCATGTAATGACGCGAGTGTGTGGCAGAAAGTATTTTGAAGCTGTAC----- 3481
DB 1747 AGCTGTGGGACGTTATGTCATGCCCATTTGTTGGCAAGGATTTTGTGATTCCTGTTCAAGCC 1806
QY 3482 ----AGCTGAGCTCTCAGCAGCGCATTCATGTTGACCTT---TTTTCAGTTTGAAGATGGA 3534
DB 1807 AAGGAGAGCGGAGGACGACGCGCTGCGCTTGCACACTAGCTACGATGATAGAGAATGCG 1866


```
QY 3535 GAATATCTTTGCCCTCTTGGCAATCTCTGTGCAATCTGTGATCCCCATTATTCCTTTG 3594
   || || || || || || || || || || || || || || || || || || || || ||
Db 1867 GAGTTCCTCTCGCCGCTCTGTGAGTCTGAGCAACACGGTATCCCC-----CTGCTG 1920

QY 3595 CAACCTCAAAAGATAAAGAGTGAAGATGCAGATGCTCTGTCTCAACTTTTGGACCTGGCA 3654
   || || || || || || || || || || || || || || || || || || || || ||
Db 1921 CTTCTCCAGGAGCATCTCGCAGGAGGTTAAATTTTTCAGACCAACAGATCTGGCA 1980

QY 3655 CGGTGGATACAGACTGTTCTGGCCAGAAATATCAGGTTATATATATAAGACATGCTAAAGGA 3714
   || || || || || || || || || || || || || || || || || || || || ||
Db 1981 CAGTGGAGCAGAGCAGTAACGCAGCAATAAAGGTGGTCCAGATGCTGAGGAGAAGCAC 2040

QY 3715 GAAACCAATTCCTATTTCTTAAATCAAGGAATGGAGATTCTACTTTGGAGTTCCAT 3774
   || || || || || || || || || || || || || || || || || || || || ||
Db 2041 AGTCTGCTGACAGCTCTTCTCAGAGGACACAGAAGCCATGAATATAATACCGATCCC 2100

QY 3775 TCCATCTGAGTTTTGGCGTTGAGTCTTCGATTAATATCAAAATAGCATCAAGAAATG 3834
   || || || || || || || || || || || || || || || || || || || || ||
Db 2101 GAAGCTTCAGGCTGATTTTATCTTAGAACCCATATTTCTGATAGCATAAAGAAATG 2160

QY 3835 GTTATCTCTTGGCCACAAATTTATAGAAATGGATTGAAGTGCACCTGATGAAGG 3894
   || || || || || || || || || || || || || || || || || || || || ||
Db 2161 TTAACGACATTTGGAAGGCTGCTTACAAGGTGGGACTGAAGGTTCACTTAATGAAGT 2220

QY 3895 GATCTCGAGTCCCATCTGACCTGGAGCACCCTGCGCTTTCACTATCCAGCAATTGAA 3954
   || || || || || || || || || || || || || || || || || || || || ||
Db 2221 GACCCCGTGGCCATCTCTGTGTGGGGACCTGTGCATACACCATCAGACGACATGAA 2280

QY 3955 AATCTATTGGAGATGAAGGAAACCTCTGTTGGAGCACCTTCAAAATAGGCAGCATAT 4014
   || || || || || || || || || || || || || || || || || || || || ||
Db 2281 AGAATTTGAGTGATGAGGAGAGCCCTGTTTGGACCTCTGCTTGTAGACTGGACAC 2340

QY 4015 GGTCTGAAAGCAATTAATGCAGTTGCGAGTTGCACAGAGGATTACCTGTCCTCAGTCC 4074
   || || || || || || || || || || || || || || || || || || || || ||
Db 2341 TGTCTCAGTCTGTTAAACAGCTTTGACAGCACATTTGGACAGTGGCGCTTACTTCTGTG 2400

QY 4075 ATACAGAAACATCTGGTTCGCTCTATCAGTTGTTCTTCCTACATATAAATCAGAAAT 4134
   || || || || || || || || || || || || || || || || || || || || ||
Db 2401 GTACAGGACACTCTCTGTAAACTCTTTGCACTCTGTTGCTAGTGACAGCTATGAAGC 2460

QY 4135 ACACCATGCTCTGTCTATAGATCTGTTTCATGTTTTGGTGGTGTGTTAGCATTC 4194
   || || || || || || || || || || || || || || || || || || || || ||
Db 2461 CTCCTGTCATACATAGACATCGATGTTTCACCTGCTGGTGGGCGCTGCTGCTTTC 2520

QY 4195 CCATCTTGTATTGGGATGACCCCTGTTGATCTGCAGCCTCTTTCAGTTAGTTCTTCCAT 4254
   || || || || || || || || || || || || || || || || || || || || ||
Db 2521 CCAGCTCTGCAGTG-----TCAGGATTTTTCAGGNAGCAGCTGCCACT 2565

QY 4255 AACCACTTTATCTCTCCATTTGATCACCATGGCACACATGCTTCAGATCTACTTACA 4314
   || || || || || || || || || || || || || || || || || || || || ||
Db 2566 GGAGACCTGCACATCTTCCACCTGGTTACCATGGCACACATCGTACAGATCTTACTTACC 2625

QY 4315 GTAGACAC---AGGCTACCCCTTGCCTCAGGTTCAAGAGACAGTGAAGGCCTATGCC 4371
   || || || || || || || || || || || || || || || || || || || || ||
Db 2626 TCATGTACAGAAGAGAAATGGCATGGATCAAGAGAAATCCCATGGGGAAGAGAACTGCC 2685

QY 4372 GCATCTCTTTCTTTTGGAGAAATTTCTCAATATACAAGTGCCTCATTTGGTGTGATATT 4431
   || || || || || || || || || || || || || || || || || || || || ||
Db 2686 ATTCCTCTTTTGCAAAACACTTCCACCATGATCTGGAAGTGCCTTGAAGAAGGCCCT 2745

QY 4432 CCTGGCTGGTATTGTGGGTCTCACTGAAGAATGGCATGCCCTTATCTTCGCTGTGCT 4491
   || || || || || || || || || || || || || || || || || || || || ||
Db 2746 CC-GCTGGCACCTGTGAGAGGCGCCGCGCCCATCATGCTTCTTCTCAAGTGCCT 2804

QY 4492 GCATGTTTTCCTACTATTTTACTTGGGTAACTCCGCTCAGGAACTGCATACCAATTTCT 4551
   || || || || || || || || || || || || || || || || || || || || ||
Db 2805 GCTTTGTTTTCCTACTATTTAAATGAGTCCCGGCCCTCCAGACCTTCAA---GTTTCT 2861

QY 4552 GCAGAAGGAGTACAGTGCCTCTGTAGCTATCTATCTTTACCTACAAATTTGTCCTG 4611
   || || || || || || || || || || || || || || || || || || || || ||
Db 2862 GGAACAAGCCATTTTGAACACTTATGTAACCTATACCTTTTCCCTACCAACCACTCATTCAC 2921

QY 4612 CTCCTCCAGGAATATTGGGACTGTGAAGGCCCTTGTCTCCAGAGGCGGTGTCAGATCCT 4671
```

```
Db 2922 CTTTTTCAAGAAAACAGTGCATCATGAACTCCTGATGTAAGAGTTGGTGCCAGAACGT 2981
   || || || || || || || || || || || || || || || || || || || || ||
QY 4672 GCCTTACTAACTGTTTGAAGCAAAAAACACCGTGTGCTCAGGTACCTTAGAAAAAGAAAT 4731
   || || || || || || || || || || || || || || || || || || || || ||
Db 2982 GAAGTTAAACGGTATCTAAATGCGGAGAGGAGCGATAGCTACCCAGAGGAGCTAAC 3041
   || || || || || || || || || || || || || || || || || || || || ||
QY 4732 AGTTTGATAGAGTTCCTGATGACTATAGTCTGCTCCTGCTCAATCAAGTTTCTCATTTTCAGG 4791
   || || || || || || || || || || || || || || || || || || || || ||
Db 3042 AAAGTATAGAGCTTCCAGAGGATTACAGCAGCTCATTAACCAAGCATCCAATTTCTCG 3101
   || || || || || || || || || || || || || || || || || || || || ||
QY 4792 TCGCCACGCTCTGCAGATGATGAGCAAGAGCATCTCTGCTCGCTTCTTCTGCGGCT 4851
   || || || || || || || || || || || || || || || || || || || || ||
Db 3102 TCGCCCAATCAGGTGGCGACAGAGCAGCTCTACTCTGTGCTCGTGTGGGAGT 3161
   || || || || || || || || || || || || || || || || || || || || ||
QY 4852 ATACTATGTTCTCAGAACATTTGCTGCCAGGAAATTTGTGAACGGGAAGAGTTGGAGCT 4911
   || || || || || || || || || || || || || || || || || || || || ||
Db 3162 CTCCTCTGCTCTCAGAGTTACTGCTGCCAAGCTGAGCTGGAGGCTGAGACGTCGAGCC 3221
   || || || || || || || || || || || || || || || || || || || || ||
QY 4912 TGCATTTTTCAGCAGCTTTCACCTGTGGAGCGGAGTCTGCAATTTTCTTAAAAATCAGAGAA 4971
   || || || || || || || || || || || || || || || || || || || || ||
Db 3222 TGCACAGCACACACCTTACTCTGTGCTCGGGGCGGCGATCTTCTGAGAGTTCCGGAA 3281
   || || || || || || || || || || || || || || || || || || || || ||
QY 4972 TCGCAGTGGTCTGTTGAAGTAAAGCCAGAGGCTGTGCCCTATCCAGCTCCTTACTTG 5031
   || || || || || || || || || || || || || || || || || || || || ||
Db 3282 TGTCAAGTGTCTATTTTAGTGGCAAAACCAAGAGATGTTTTTATTCTCTCTTACCTT 3341
   || || || || || || || || || || || || || || || || || || || || ||
QY 5032 GATGAATATGGAGAAACAGACCTTGGCTTGAAGAGGGCAACCCCTTCAATTTATCTCGT 5091
   || || || || || || || || || || || || || || || || || || || || ||
Db 3342 GACGACTATGGGAGACCCAGCAGGACTCAGCAGAGGAATCCTTTACATTTATGCCAA 3401
   || || || || || || || || || || || || || || || || || || || || ||
QY 5092 GAGCGGTATCGGAAGCTCCATTTGGTCTGCAACACACTGCAATATAGAAGAGATTGCT 5151
   || || || || || || || || || || || || || || || || || || || || ||
Db 3402 GAGCGTTCGAAAGATCCAGAAGCTCTGSCAGCAGCATAGTATCACAGAGGAGATCGGA 3461
   || || || || || || || || || || || || || || || || || || || || ||
QY 5152 AGAGCCACAGACACTAATCAGATGTTATTGATTCACTGCGAG 5196
   || || || || || || || || || || || || || || || || || || || || ||
Db 3462 CACGCGCAGGAGGCTAACACAGACCCCTGTCGGAATTTGACTGGCAG 3506
   || || || || || || || || || || || || || || || || || || || || ||

RESULT 4
BQ233617
LOCUS
DEFINITION
BQ233617
VERSION
BQ233617.1 GI:20415017
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 865)
AUTHORS
NIH-MGC http://mgc.ncl.nih.gov/.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: ARCC/DCRP/DTP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13294 row: 1 column: 04
High quality sequence stop: 720.
Location/Qualifiers
1..865
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6047379"

FEATURES
source
```

```
/tissue_type="melanotic melanoma"
/libb_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_72"
/notes="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2 kb. Library constructed by Life
Technologies."
BASE COUNT      261 a   146 c   204 g   250 t     4 others
ORIGIN
Query Match      14.8%; Score 771.4; DB 13; Length 865;
Best Local Similarity 99.0%; Pred. No. 1e-147;
Matches 786; Conservative 0; Mismatches 7; Indels 1; Gaps 1;
QY 1601 TCCAAGAGTGGTGTGTCGATGACGAACCTCTTACTTGGCTTATAAAGAAATGTACAC 1660
Db      |||
8 TCCGAGAGTGGTGTGTCGATGACGAACCTCTTACTTGGCTTATAAAGAAATGTACAC 67
QY 1661 AAGCTGTGATGAGTGCAGTACCAGTTTCATATCTAGTAGCAACAGTAGTACAATCGT 1720
Db      |||
68 AAGCTGTGATGAGTGCAGTACCAGTTTCATATCTAGTAGCAACAGTAGTACAATCGT 127
QY 1721 GTGGACATAGTTTGGAAACAAAGTCCCTACAGATATCTGAGGATCTTTGAAGCATACATC 1780
Db      |||
128 GTGGACATAGTTTGGAAACAAAGTCCCTACAGATATCTGAGGATCTTTGAAGCATACATC 187
QY 1781 TGGCACTCTAGACCCCTTGGTGGTCTTCATGATGATGTTAAAGCAGGCTGGTGTCTGTT 1840
Db      |||
188 TGGCACTCTAGACCCCTTGGTGGTCTTCATGATGATGTTAAAGCAGGCTGGTGTCTGTT 247
QY 1841 CAAGACTGCATGAATTTGTCCTTTTTCAGGACCTTCAAGTAGAGGTACTAGTGAATATC 1900
Db      |||
248 CAAGACTGCATGAATTTGTCCTTTTTCAGGACCTTCAAGTAGAGGTACTAGTGAATATC 307
QY 1901 CTTTACGTTCTGCTGGTGTGCCAGGTTGTCGTGAGATGTGGCGAAGAAATGGAC 1960
Db      |||
308 CTTTACGTTCTGCTGGTGTGCCAGGTTGTCGTGAGATGTGGCGAAGAAATGGAC 367
QY 1961 TGTCTCTTATAGCCAGGTGTTTATTAACCAAGATGTTAAGTGCAGAGAAATGTATG 2020
Db      |||
368 TGTCTCTTATAGCCAGGTGTTTATTAACCAAGATGTTAAGTGCAGAGAAATGTATG 427
QY 2021 ATAAAGATATCATGCTTCAGATGGTGCATCTTTAATGGATCCCAATAAGTCTTGT 2080
Db      |||
428 ATAAAGATATCATGCTTCAGATGGTGCATCTTTAATGGATCCCAATAAGTCTTGT 487
QY 2081 TACTGTCTTACAGAGTATGAACCTTCCCGAGGCTTTTAAACAGACCATATCTACAAAG 2140
Db      |||
488 TACTGTCTTACAGAGTATGAACCTTCCCGAGGCTTTTAAACAGACCATATCTACAAAG 547
QY 2141 ACCAGGATTTGATTAAACAATATAATACACTAATAGAGAAATGCTTCAGTCCCTCATCT 2200
Db      |||
548 ACCAGGATTTGATTAAACAATATAATACACTAATAGAGAAATGCTTCAGTCCCTCATCT 607
QY 2201 ATATTGGGTGAGCGTTATGTACCTGGAGTGGGAATGTGACCAAGAGAGGTACAA 2260
Db      |||
608 ATATTGGGTGAGCGTTATGTACCTGGAGTGGGAATGTGACCAAGAGAGGTACAA 667
QY 2261 TGAGAGAAATCATTCACCTTGTTCATTGACCCATGCCACAGTGGCCATTGCCAAA 2320
Db      |||
668 TGAGAGAAATCATTCACCTTGTTCATTGACCCATGCCACAGTGGCCATTGCCAAA 727
QY 2321 ATTTACCTGAGATGAAATAATGAAACT - GGCTTACAGATGTGCATAAACAAAGTGGCC 2379
Db      |||
728 ATTTACCTGAGATGAAATAATGAAACTGGGCTTANAGATGTGCATAAACAAAGTGGCC 787
QY 2380 ACATTTAAGAAACC 2393
Db      |||
788 ACATTTAAGAAAC 801
RESULT 5
BX104087
```

LOCUS BX104087 712 bp mRNA linear EST 06-FEB-2003
DEFINITION Soares_testis_NHT Homo sapiens cDNA clone IMAGE998F081824
; IMAGE:742663, mRNA sequence.
ACCESSION BX104087
VERSION BX104087.1 GI:27846028
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 712)
AUTHORS Ebert,L., Heil,O., Hennig,S., Neubert,P., Partsch,E., Peters,M.,
Radelof,U., Schneider,D. and Korn,B.
TITLE Human Unigeneset - RZPD3
JOURNAL Unpublished
COMMENT Contact: Ina Rolfes
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany
RZPD: IMAGP998F081824.
RZPDLIB: I.M.A.G.E. cDNA Clone Collection;
Human Unigeneset - RZPD3 (RZPDLIB No.972)
http://www.rzpd.de/CloneCards/cgi-
bin/showLib.pl.cgi/response?libNo=972 Contact: Ina Rolfes
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Heubnerweg 6, D-14059 Berlin, Germany
Tel: +49 30 32639 101
Fax: +49 30 32639 111
www.rzpd.de
This clone is available royalty-free from RZPD;
contact RZPD (clone@rzpd.de) for further information. Seq primer:
M13r, Primer sequence: TTTACACAGGAAACAGCTATGAC.
FEATURES
Location/Qualifiers
1..712
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGP998F081824 ; IMAGE:742663"
/sex="male"
/lab_host="DH10B"
/clone_lib="Soares_testis_NHT"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech Laboratories
, Inc., and primed with a Not I - oligo(dt) primer [5',
TCTTACCAATCTGAATGGGAGCGGCCCAATTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 174 a 164 c 173 g 201 t
ORIGIN
Query Match 13.6%; Score 708.8; DB 13; Length 712;
Best Local Similarity 99.7%; Pred. No. 7.1e-135;
Matches 710; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 4433 CTGGCTGGTATTGTGGGTCTCAGTGAAGAATGGCATCACCCCTTATCTTCGTGTGCTG 4492
Db 1 CTGGCTGGTATTGTGGGTCTCAGTGAAGAATGGCATCACCCCTTATCTTCGTGTGCTG 60
QY 4493 CATTTGTTTTCCACTATTACTTGGGTAACCTCCGCTGAGGAACATGCATACCAATTCG 4552
Db 61 CATTTGTTTTCCACTATTACTTGGGTAACCTCCGCTGAGGAACATGCATACCAATTCG 120
QY 4553 CAGAGGAGAGTACAGTGCATCTGTAGTATCTATCTTTTACCTACAAATTTGTTCTGTC 4612
Db 121 CAGAGGAGAGTACAGTGCATCTGTAGTATCTATCTTTTACCTACCAATTTGTTCTGTC 180
QY 4613 TCTTCCAGGAATATTGGGATAGTGTAAAGCCCTTGTCTCCAGAGCGGTGTGCAGATCCCTG 4672
Db 181 TCTTCCAGGAATATTGGGATAGTGTAAAGCCCTTGTCTCCAGAGCGGTGTGCAGATCCCTG 240

```
QY 4673 CTTTACTAACTGTTTGAAGCAAAAAACACCGTGGTGCAGGTACCTAGAGAAAAAATA 4732
Db 241 CTTTACTAACTGTTTGAAGCAAAAAACACCGTGGTGCAGGTACCTAGAGAAAAAATA 300
QY 4733 GTTTCATAGAGTCTCTGATGACTATAGTGCCTCCTGAATCAAGTCTCTATTCAGGT 4792
Db 301 GTTTCATAGAGTCTCTGATGACTATAGTGCCTCCTGAATCAAGTCTCTATTCAGGT 360
QY 4793 GCCACGCTGTCAGATGATGAGCAAAAGCATCTCTCTGTCGCCCTTTCTGTGGGCTA 4852
Db 361 GCCACGCTGTCAGATGATGAGCAAAAGCATCTCTCTGTCGCCCTTTCTGTGGGCTA 420
QY 4853 TACTATGTTCTCAGAACATTTGCTGCCAGAAATTTGTGAACGGGAAGAGTTGGAGCTT 4912
Db 421 TACTATGTTCTCAGAACATTTGCTGCCAGAAATTTGTGAACGGGAAGAGTTGGAGCTT 480
QY 4913 GCATTTTTCAGCACTTCACCTGTGGAGCGGAGTCTGCATTTTCTTAAATCAGAGAT 4972
Db 481 GCATTTTTCAGCACTTCACCTGTGGAGCGGAGTCTGCATTTTCTTAAATCAGAGAT 540
QY 4973 GCCGAGTGGTCTGTTGAAAGTAAAGCCAGAGGCTGTCCTATCCAGCTCCTTACTTGG 5032
Db 541 GCCGAGTGGTCTGTTGAAAGTAAAGCCAGAGGCTGTCCTATCCAGCTCCTTACTTGG 600
QY 5033 ATGAATATGGAGAAACAGACCTGCGCTGAAGAGGGCAACCCCTTCTATTTATCTCGTG 5092
Db 601 ATGAATATGGAGAAACAGACCTGCGCTGAAGAGGGCAACCCCTTCTATTTATCTCGTG 660
QY 5093 AGCGGTATCGGAAGCTCCATTTGTTGTCGCAACACACTGCATTTATAGAAGA 5144
Db 661 AGCGGTATCGGAAGCTCCATTTGTTGTCGCAACACACTGCATTTATAGAAGA 712

RESULT 6
LOCUS AI929033
DEFINITION au64c10 y1 Schneider fetal brain 00004 Homo sapiens cDNA clone
IMAGE:2519538 5' similar to TR:070481 070481 UBIQUITIN-PROTEIN
LIGASE E3 COMPONEN N-RECOGNIN ;, mRNA sequence.
ACCESSION AI929033
VERSION AI929033.1 GI:5664997
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 682)
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin
J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,
White, Y., Wylie, T., Waterston, R. and Wilson, R.
WashU-NCI human EST Project
Unpublished
Other_ESTs: au64c10.x1
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Possible reversed clone: similarity on wrong strand
Seq primer: -40RP from Gibco
High quality sequence stop: 450.
location/Qualifiers
1..682
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2519538"
/sex="male"
/tissue_type="frontal lobe"
```

```
/dev_stage="5 months post-conception"
/lab_host="DH10B"
/clone_lib="Schneider fetal brain 00004"
/notes="Organ: brain; Vector: pBluescript SK (Stratagene);
Site_1: SstI; Site_2: XhoI; Double-stranded cDNA was
prepared from human fetal brain tissue. 5' and 3'
adaptors were used in cloning as follows: 5' adaptor
sequence:
5'-GAGAGAGAGAGAGCTCAAGATCCTTAATAATATCCCCCCCCCC-3'
and 3' adaptor sequence:
5'-GAGAGAGAGAGACTCGAGTTTTTTTTTTTTTT-3'. The library was
size-selected for >0.5 kb inserts and has an average
insert size estimated at 1.2 kb. This library was
constructed using the CAP-trapper method for full-length
enrichment and has not undergone amplification. Library
was constructed by Dr. Claudio Schneider (LNCIB-Area
Science Park, Trieste, Italy). "
BASE COUNT 178 a 155 c 166 g 182 t 1 others
Query Match 12.9%; Score 673; DB 9; Length 682;
Best Local Similarity 99.1%; Pred. No. 1.6e-127;
Matches 676; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 4521 AACTCCGCTGAGGAACACTGCATACCAATTCGCAGAGGAGAGTACAGTGCACCTCTAG 4580
Db 1 AACTCCGCTGAGGAACACTGCATACCAATTCGCAGAGGAGAGTACAGTGCACCTCTAG 60
QY 4581 CTATCTATCTTTACCTACAAATTTGTTCTGCTCTCCAGCAATATTTGGGATCTGTAAG 4640
Db 61 CTATCTATCTTTACCTACAAATTTGTTCTGCTCTCCAGCAATATTTGGGATCTGTAAG 120
QY 4641 GCCCTTGTCCAGAGCGGTGTGCAGATCTCGCTTACTAACTGTTTGAAGCAAAAAA 4700
Db 121 GCCCTTGTCCAGAGCGGTGTGCAGATCTCGCTTACTAACTGTTTGAAGCAAAAAA 180
QY 4701 CACCGTGTGTCAGTACCCCTAGAAAAAGAAATATTTGATAGAGCTTCCTGATGACTATAG 4760
Db 181 CACCGTGTGTCAGTACCCCTAGAAAAAGAAATATTTGATAGAGCTTCCTGATGACTATAG 240
QY 4761 CTGCTCTCTGAATCAAGCTTCTCATTTTCAGGTGCCCGCTGCGAGATCATGCGCAA 4820
Db 241 CTGCTCTCTGAATCAAGCTTCTCATTTTCAGGTGCCCGCTGCGAGATCATGCGCAA 300
QY 4821 GCATCTGTCTCTGCTCTTCTTGTGGGCTATCTATGTTCTCAGAACATTTGCTGCCA 4880
Db 301 GCATCTGTCTCTGCTCTTCTTGTGGGCTATCTATGTTCTCAGAACATTTGCTGCCA 360
QY 4881 GGAATTTGAGCGGGAAGAGGTTGGAGCTTGCATTTTTCACGCACTTCACCTGTGAGC 4940
Db 361 GGAATTTGAGCGGGAAGAGGTTGGAGCTTGCATTTTTCACGCACTTCACCTGTGAGC 420
QY 4941 CGGAGTCTGCATTTTCTTAAATCAGAGATCCGAGTGGTCTGTTGAAGTAAAGC 5000
Db 421 CGGAGTCTGCATTTTCTTAAATCAGAGATCCGAGTGGTCTGTTGAAGTAAAGC 480
QY 5001 CAGAGCTCTGCTATCCAGCTCCTTACTTGGATGAATATGGAGAAACAGACCTGSCCT 5060
Db 481 CAGAGCTCTGCTATCCAGCTCCTTACTTGGATGAATATGGAGAAACAGACCTGSCCT 540
QY 5061 GAAGGGGCAACCCCTTCTTATTTATCTGTAGCGGTATCGGAAGCTCCATTTGGTCTG 5120
Db 541 GAAGGGGCAACCCCTTCTTATTTATCTGTAGCGGTATCGGAAGCTCCATTTGGTCTG 600
QY 5121 GCAACAACTGCATTTATAGAGAGATTCGTAGAGCCAGAGACTTAATCAGATGTTATT 5180
Db 601 GCAACAACTGCATTTATAGAGAGATTCGTAGAGCCAGAGACTTAATCAGATGTTATT 660
QY 5181 TGGATTCAACTGGCAGTTACTG 5202
Db 661 GGGATTCACTCCCTGAGTACTG 682
```

```
RESULT 7
BU618516/c
LOCUS      668 bp      mRNA      linear      EST 23-SEP-2002
DEFINITION UI-H-FH1-bfk-c-21-0-UI.s1 NCI_CGAP_FH1 Homo sapiens cDNA clone
            UI-H-FH1-bfk-c-21-0-UI 3', mRNA sequence.
ACCESSION  BU618516
VERSION     BU618516.1 GI:23284731
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 668)
REFERENCE  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
            National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
            Unpublished
            Contact: Robert Strausberg, Ph.D.
            Email: cgaps-remail.nih.gov
            Tissue Procurement: James Martin
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Clone distribution information can be obtained
            from Dr. M. Bento Soares, bento-soares@uiowa.edu
            Seq primer: M13 FORWARD
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source           1..668
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UI-H-FH1-bfk-c-21-0-UI"
                     /tissue_type="Cell Line"
                     /dev_stage="Adult"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="NCI_CGAP_FH1"
                     /note="Organ: Chondrosarcoma; Vector: pT7T3-Pac (Pharmacia
                     ) with a modified polylinker; Site_1: Ecor I; Site_2: Not
                     I; NCI_CGAP_FH1 is a normalized cDNA library obtained from
                     a cell line derived from grade I chondrosarcoma tissue.
                     The library was constructed and normalized according to
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. First strand cDNA synthesis was primed with an
                     oligo-dT primer containing a Not I site. Double stranded
                     cDNA was ligated to an Ecor I adaptor, digested with Not
                     I, and cloned directionally into pT7T3-Pac vector. The
                     oligonucleotide used to prime the synthesis of
                     first-strand cDNA contains a library tag sequence that is
                     located between the Not I site and the (dT)18 tail. The
                     sequence tag for this library is AGAATCCGGC. The cell line
                     was provided by Dr. James Martin from the University of
                     Iowa.
                     TAG_LIB=UI-H-FH1
                     TAG_TISSUE=Human Chondrosarcoma Cell Line CS8 - Grade 1
                     Chondrosarcoma
                     TAG_SEQ=AGAATCCGGC"
BASE COUNT  180 a 143 c 110 g 235 t
ORIGIN
Query Match      12.8%; Score 664.8; DB 13; Length 668;
Best Local Similarity 99.7%; Pred. No. 7.7e-126;
Matches 666; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1867 GAGGACTTCAAGTAGAGGTACTAGTGGATATCCTTACGTTGTCGTGTTGGTTGCC 1926
      |||||||
Db 668 GAGGACTTCAAGTAGAGGTACTAGTGGATATCCTTACGTTGTCGTGTTGGTTGCC 609
      |||||||
QY 1927 CAGGTTGTTGCTCAGATGTGGCGAAGAAATGGAGTGTCTTATTAGCCAGGTGTTTAT 1986
      |||||||
Db 608 CAGGTTGTTGCTCAGATGTGGCGAAGAAATGGAGTGTCTTATTAGCCAGGTGTTTAT 549
      |||||||
QY 1987 TACCAAGATGTTAAGTGCAGAGAGAAATGTATGATAAAGATATCATCATGCTTCAGATT 2046
```

```
Db 548 TACCAAGATGTTAAGTGCAGAGAGAAATGTATGATAAAGATATCATCATGCTTCAGATT 489
      |||||||
QY 2047 GGTGCATCTTTAATGGATCCCAATAAGTTCCTTGTACTGTTACTTCCAGAGGTATGAACCT 2106
      |||||||
Db 488 GGTGCATCTTTAATGGATCCCAATAAGTTCCTTGTACTGTTACTTCCAGAGGTATGAACCT 429
      |||||||
QY 2107 GCCGAGGCTTTTAAACAAGACCATATCTACAAAAGACCAGGATTTGATTAACAATATAAT 2166
      |||||||
Db 428 GCCGAGGCTTTTAAACAAGACCATATCTACAAAAGACCAGGATTTGATTAACAATATAAT 369
      |||||||
QY 2167 ACACATAATAGAGAAATGCTTCAGGTCCTCATCTATATTGTTGGGTGAGGTTATGTACCT 2226
      |||||||
Db 368 ACACATAATAGAGAAATGCTTCAGGTCCTCATCTATATTGTTGGGTGAGGTTATGTACCT 309
      |||||||
QY 2227 GGAGTGGGAAATGTGACCAAAAGAGGTCACAATGAGAGAAATCATTCACCTGCTTTTGC 2286
      |||||||
Db 308 GGAGTGGGAAATGTGACCAAAAGAGGTCACAATGAGAGAAATCATTCACCTGCTTTTGC 249
      |||||||
QY 2287 ATTGAACCCATGCCACACAGTGCCTATGCAAAAATTTTACCTGAGAAATGAAAATAATGAA 2346
      |||||||
Db 248 ATTGAACCCATGCCACACAGTGCCTATGCAAAAATTTTACCTGAGAAATGAAAATAATGAA 189
      |||||||
QY 2347 ACTGCTTTAGAGATGTCTATAAACAAGTGGCCACATTTAAGAAACCCAGGTGTATCAGGC 2406
      |||||||
Db 188 ACTGCTTTAGAGATGTCTATAAACAAGTGGCCACATTTAAGAAACCCAGGTGTATCAGGC 129
      |||||||
QY 2407 CATGGAGTTTATGAACATAAAGATGAATCAGTCAGAAAGACTTCAATATGTACTTTTATCAT 2466
      |||||||
Db 128 CATGGAGTTTATGAACATAAAGATGAATCAGTCAGAAAGACTTCAATATGTACTTTTATCAT 69
      |||||||
QY 2467 TACTCCAAAACCCAGCATAGCAAGGCTGAACATATGCAAGAAAGAGAGAAACCAAGAA 2526
      |||||||
Db 68 TACTCCAAAACCCAGCATAGCAAGGCTGAACATATGCAAGAAAGAGAGAGAAACCAAGAA 9
      |||||||
QY 2527 AACAAAGA 2534
      |||||||
Db 8 AACAAAAA 1

RESULT 8
BU6534574
LOCUS      756 bp      mRNA      linear      EST 03-APR-2001
DEFINITION 602553425F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4663182 5',
            mRNA sequence.
ACCESSION  BU6534574
VERSION     BU6534574.1 GI:13526116
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 756)
REFERENCE  NIH-MGC http://mgc.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished
            Contact: Robert Strausberg, Ph.D.
            Email: cgaps-remail.nih.gov
            Tissue Procurement: CLONTECH Laboratories, Inc.
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCMI465 row: a column: 07
            High quality sequence stop: 751.
FEATURES             Location/Qualifiers
     source           1..756
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4663182"
```

lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_77"
/note="Organ: lung; Vector: pDNR-LIB (Clontech); Site_1: SfiI (ggcccctcgcc); Site_2: SfiI (ggccattggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCATTATGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGCGGCACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.9 kb (range 0.5-4.0 kb). 12/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."
BASE COUNT 274 a 138 c 152 g 192 t
ORIGIN

Query Match	12.6%	Score	656.6	DB	10	Length	756
Best Local Similarity	95.9%	Pred. No.	3.8e-124				
Matches	730	Conservative	0	Mismatches	19	Indels	12
				Gaps	5		
QY	2096	GGTATGAACCTGCCGAGCTTTTACAAAGACCATATCTACAAAAGACCAGGATTTGATTA	2155				
Db	1	GGTATGAACCTGCCGAGCTTTTACAAAGACCATATCTACAAAAGACCAGGATTTGATTA	60				
QY	2156	AACATATATACACTATAGAAGAAATGCTTCAGGTCTCATCTATATTTGGGTGAGC	2215				
Db	61	AACATATATACACTATAGAAGAAATGCTTCAGGTCTCATCTATATTTGGGTGAGC	120				
QY	2216	GTTATGTACCTGGAGTGGAAATGTGACCAAGAGAGGTGCACATGAGAGAAATCATTC	2275				
Db	121	GTTATGTACCTGGAGTGGAAATGTGACCAAGAGAGGTGCACATGAGAGAAATCATTC	180				
QY	2276	ACTTGTCTTGGATTGAACCCATGCCACAGTGCATTCGCCAAAATTTACCTGAGAATG	2335				
Db	181	ACTTGTCTTGGATTGAACCCATGCCACAGTGCATTCGCCAAAATTTACCTGAGAATG	240				
QY	2336	AAAATAATGAACCTGGCTTAGAGATGTCTATAAACAAGTGGCCACATTTAAGAACACAG	2395				
Db	241	AAAATAATGAACCTGGCTTAGAGATGTCTATAAACAAGTGGCCACATTTAAGAACACAG	300				
QY	2396	GTGTATCAGGCGCATGGAGTTTATGAACCTAAAGATGAATCACTGAAAGACCTTCAATATGT	2455				
Db	301	GTGTATCAGGCGCATGGAGTTTATGAACCTAAAGATGAATCACTGAAAGACCTTCAATATGT	360				
QY	2456	ACTTTTATCATTTCTCCAAACCCAGATAGCAGGCTGACATATGACAGAGAAAGGA	2515				
Db	361	ACTTTTATCATTTCTCCAAACCCAGATAGCAGGCTGACATATGACAGAGAAAGGA	420				
QY	2516	GAACAACAGAAACAAAGATGAAG-----CATTCGCCGCCACCCACCTCTCGAATCT	2569				
Db	421	GAACAACAGAAACAAAGATGAAGTAAACATTTGCCGCCACCCACCTCTCGAATCT	480				
QY	2570	GCCCTGCTTTCAGCAAGTGATTAACCTTCTCAACTGTGATAT-CATGATGTACATTTCTC	2628				
Db	481	GCCCTGCTTTCAGCAAGTGATTAACCTTCTCAACTGTGATATCCATGATGTACATTTCTC	540				
QY	2629	AGGACCGTATTGACGGCGCAATAGACACAGATTTACTTGTGGACCGAGGGATGCTC	2688				
Db	541	AGGACCGTATTGACGGCGCAATAGACACAGATTTACTTGTGGACCGAGGGATGCTC	600				
QY	2689	CAATGGCTTTTCATATTTCTGCATTTGGTGTCTACTAGAGAGAACACAGCTTCAAAA	2748				
Db	601	CAATGGCTTTTCATATTTCTGCATTTGGTGTCTACTAGAGAGAACACAGCTTCAAAA	659				
QY	2749	GCTCCTGAAGAAGAGTAAACATTTTACATTAAGGCTTCAAGATTTGGAAGTTTCA	2808				
Db	660	GCTCCTGAAGAAGAGTAAACATTTTACATTAAGGCTTCAAGATTTGGAAGTTTCA	719				
QY	2809	GCCATGAATATACAAATGCTTTTGGAAAACTCAAAAGAAAT	2849				
Db	720	G-CATGAATATACAA--TGTTTTGAAAAATCAAAAGAAAT	756				

RESULT 9

AK083320	3627 bp	mRNA	linear	HTC 05-DEC-2002
LOCUS				
DEFINITION	Mus musculus 2 days neonate thymus cells cDNA, RIKEN full-length enriched library, clone: C920004H05 product: similar to (A PROTEIN SIMILAR TO UBIQUITIN-PROTEIN LIGASE E3 COMPONENT N-RECOGNIN (UBIQUITIN-PROTEIN LIGASE E3-ALPHA)) (FRAGMENT) (Homo sapiens), full insert sequence.			
ACCESSION	AK083320			
VERSION	AK083320.1	GI:26350448		
KEYWORDS	HTC; CAP trapper.			
SOURCE	Mus musculus (house mouse)			
ORGANISM	Mus musculus			
REFERENCE	1 Carninci, P. and Hayashizaki, Y. High-efficiency full-length cDNA cloning Meth. Enzymol. 303, 19-44 (1999) 99279253 10349636			
AUTHORS	2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000) 20499374 11042159			
TITLE	3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Katsunari, T., Tashiro, H., Itoh, A., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohata, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000) 20530913 11076861			
JOURNAL	4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gotohori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaide, I., Pesole, G., Tomita, M., Quackenbush, J., Schriml, L. M., Staib, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyo-oka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohtsuki, S. and Hayashizaki, Y. Functional annotation of a full-length mouse cDNA collection Nature 409 (6821), 685-690 (2001) 21085660 11217851			
AUTHORS	5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002) 6 (bases 1 to 3627) Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,			

Query Match		11.9%;	Score 620;	DB 12;	Length 797;
Best Local Similarity		88.3%;	Pred. No. 1.3e-116;		
Matches 685;		Conservative 0;	Mismatches 90;	Indels 1;	Gaps 1;
QY	4431	TCCTGGCTGCTATTTCTGGTCTCACTGAAGATGCAACCCCTTATCTTCGCTGTC	4490		
Db	2	TCCTGGCTGCTATTTCTGGTCTCACTGAAGATGCAACCCCTTATCTTCGCTGTC	61		
QY	4491	TGCAATGTTTTCACATATTACTTGGGTAATACTCCGCTCGAGAACTGCATACCAATTC	4550		
Db	62	TGCACTGCTTTTCCCACTATTACTTGGAGTAGTCGCGCTGGAAGAACTGTTGCCAATTC	121		
QY	4551	TCGAGAGCAGATCAGTGCATCTGTAGCTATCTATCTTACCTACCAATTTGTCTCT	4610		
Db	122	TGCTGAAGAGAAATTCAGTGCATCTGTAGCTATCTATCTTACCCACAAATTTGTCTCT	181		
QY	4611	GCTCTTCCAGGAATATTGGATACGTGAAGGCCCTTGTCCAGAGCGGTGTCAGATCC	4670		
Db	182	GCTTTTCCAGGAATATTGGATACCAATAGGCCCTTACTACAGAGGTGGTGGAGATCC	241		
QY	4671	TGCTTTACTAAATGTTTGAAGCAAAAACACCGTGGTCAGGTACCCCTAGAAAAAGAAA	4730		
Db	242	TGCTTTACTCAAGTCTTTGAAGCAGAAAGTGTGTGGTCAGGTACCCCTAGAAAAAGAAA	301		
QY	4731	TAGTTGTATAGAGTTCCTGTAGTACTATAGTGCCTCTCTGAAATCAAGCTTCTCATTTAG	4790		
Db	302	TAGTTGTATAGAGTTCCTGTAGGACTTACAGCTGTCTTCTAAATCAGGCTTCTCATTAG	361		
QY	4791	GTGCCACGGTCTGAGATGATCAGCGAAAGCATCTGTCTCTCTGCTTTCTGTGGGC	4850		
Db	362	GTGTCACGGTCTGAGATGATGAGCGAAAGCATCTGTCTCTCTGCTTTCTGTGGGC	421		
QY	4851	TATACATGTTCTCAGAACATTTGTCGCCAGAAATTTGTAACGGGAAGAGTTGGAGC	4910		
Db	422	CATCTGTGTTCTCAGACATCTGTTGCCAAGAAATAGTGAATGGGAAGAGTTGGAGC	481		
QY	4911	TTCATTTTTCACGCACTTCACTGTGGAGCGGAGTCTGCATTTTCTTAAATCAGAGA	4970		
Db	482	GTGCGTTTTCATGCGCTTCATTTGTGTGCTGTGAGTCTGCAATTTCTTAAATCCGAGA	541		
QY	4971	ATGCCAGTGTCTGTTGAAGTAAAGCCAGAGGCTGTGCTATCCAGCTCCCTTACTT	5030		
Db	542	ATGCAGGGTGTCTGTTGGAGGAAAGCCAGAGGCTGTGCTACCCAGCCCTTACTT	601		
QY	5031	GGATGAATATGGAGAAACAGACCTGGCTGAAGAGGGCAACCCCTTCAATTTATCTCG	5090		
Db	602	GGATGAATATGGAGAAACAGACCCAGGGCTTAAGAGAGGAACCCACTTCAATTTATCTCG	661		
QY	5091	TGAGCGGTATCGAAGCTCCATTTGGTC-TGGCAACAACACTGCAATATAGAGAGATTG	5149		
Db	662	GGAGCGGTATCGAAGCTCAATTTGTCTTGGCAAAAGGACTGGATTATAGAGAGAAATG	721		
QY	5150	CTAGGAGCCAAAGACTAATCAGATGTTATTTGGATTCACTGAGGAGTACTGTGA	5205		
Db	722	CTCGAGCGGGAGACTAATCAGATGCTATTTGGACTTAACCTGGGAGTACTCTGA	777		
RESULT 11					
BU955331					
LOCUS					
DEFINITION					
AGENCY					
IMAGE:6727400 5', mRNA sequence.					
ACCESSION					
BU955331					
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
Homo sapiens (human)					
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
1 (bases 1 to 774)					
NIH-MGC http://mgi.nci.nih.gov/					
National Institutes of Health, Mammalian Gene Collection (MGC)					
Unpublished					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
COMMENT					
Contact: Robert Strausberg, Ph.D.					
Email: cgaaps-r@mail.nih.gov					
Tissue Procurement: NCI					
CDNA Library Preparation: Michael Brownstein Laboratory					
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)					
Clone Distribution: MGC clone distribution information can be					
found through the I.M.A.G.E. Consortium/LLNL at:					
http://image.llnl.gov					
Plate: LLCM3047 row: f column: 07					
High quality sequence stop: 555.					
FEATURES					
source					
1..774					
/organism="Homo sapiens"					
/mol_type="mRNA"					
/db_xref="taxon:9606"					
/clone="IMAGE:6727400"					
/tissue_type="mixed (pool of 40 RNAs)"					
/lab_host="DH10B (TI-phage-resistant)"					
/clone_lib="NIH_MGC_126"					
/note="Vector: pDNR-LIB; Site 1: SfiI (ggccattatggcc);					
Site 2: SfiI (ggcgctcgcc); Double-stranded cDNA was					
prepared from a pool of 40 cell line polyA+ RNAs (bladder					
- 2%, blood - 33.4%, brain - 5.6%, breast - 12.5%, colon -					
4%, connective tissue - 1.4%, eye - 1%, intestine - 2.6%,					
kidney - 2.2%, liver - 5.7%, lung - 10.8%, NK-cell - 5.2%					
, ovary - 4%, pharynx - 2.5%, prostate - 4.3%, salivary					
gland - 1.3%, and skin - 2.3%). 5' and 3' adaptors were					
used in cloning as follows:					
5'-AAGCAGTGGTATCAACGAGTGGCCATTACGCGCGG-3' and					
5'-ATTCTAGAGCGGCGCGCATATG-DT(30)NN-3'. Full-length					
enriched library was constructed using the Clontech					
Creator SMART kit and size-selected to contain the 0.5-1					
kb size fraction (other fractions present in NIH_MGC_127					
and NIH_MGC_128). Library created in the laboratory of T.					
Ushdin, M.D., Ph.D. (NIMH, NIH). Note: this is a NIH_MGC					
Library."					
BASE COUNT					
ORIGIN					
247 a 151 c 154 g 222 t					
Query Match					
Best Local Similarity					
Matches 637; Conservative 5; Mismatches 28; Indels 6; Gaps 2;					
QY	671	HSYDHGTCATATACAGCTTACAAAGAGCTCTTGACTGTGAGCTCCAGAGGCCAGTTGC	730		
Db	39	ACCAGCTCATATACAGCTTACAAAGAGCTCTTGACTGTGAGCTCCAGAGGCCAGTTGC	98		
QY	731	ATACCACTGCCATTGCAAAAGAGGTCGTCGGGCTGTTTAAAGCGGGAGCTTATGCTGCT	790		
Db	99	ATACCACTGCCATTGCAAAAGAGGTCGTCGGGCTGTTTAAAGCGGGAGCTTATGCTGCT	158		
QY	791	GCAGGAAGCAAGGAAGATATAAGAGTCAATTAGAAAATGTCTCTCAACATCCACTTC	850		
Db	159	GCAGGAAGCAAGGAAGATATAAGAGTCAATTAGAAAATGTCTCTCAACATCCACTTC	218		
QY	851	ATGTAGAAGTATTACACTCAGAGATTATGGCTCATCAGAAATTTCTCTTGGCTCTGCTT	910		
Db	219	ATGTAGAAGTATTACACTCAGAGATTATGGCTCATCAGAAATTTCTCTTGGCTCTGCTT	278		
QY	911	CCTGGATGAACAAAATTTATGAGCTTATCAAGTGAATTTAGGAGATCTTTTGGCAAGAT	970		
Db	279	CCTGGATGAACAAAATTTATGAGCTTATCAAGTGAATTTAGGAGATCTTTTGGCAAGAT	338		
QY	971	GCCTTAGAGAAGAACCTGACTCGGAGATCCCTGTCTCATAGCAGGTTAATGCTTTGGG	1030		
Db	339	GCCTTAGAGAAGAACCTGACTCGGAGATCCCTGTCTCATAGCAGGTTAATGCTTTGGG	398		
QY	1031	ATGCAAGCTTTATAAGGTCGCCGTAGATCCCTCATGAATTTGATCTTCAGCAGCTTTT	1090		
Db	399	ATGCAAGCTTTATAAGGTCGCCGTAGATCCCTCATGAATTTGATCTTCAGCAGCTTTT	458		
QY	1091	TTATGGAGATGGAATACAAAAAATCTTTTGCTATGGAATTTGTGAAGTATTATAACAAC	1150		

```

|||||
459 TTATGGAGTGAATACAAAACCTCTTGGTATGGAAATTTGTGAAGTATTTATTAACAC 518
QY TGCAGAAAGATATATCATGATGATCATGACAGAAATATCTCTATTAACATCAGTTTCAG 1210
Db TGCAGAAAGATATATCATGATGATCATGACAGAAATATCTCTATTAACATCAGTTTCG 578
QY TTTCAGATGTTTACTGTTCTTCTACTCTGGTCGACATCTTATTTGAAGACGAGAAATGTTATCT 1270
Db TTTCAGATGTTTACTGTTCTTCTACTCTGGTCGACATC---TTGAAAACAGAAAGTTATCC 635
QY CTGTCATTACTGAACACTCTGCTAGAA--GTTTACCTGAGTACTTGGACAGGACAAATA 1327
Db CTGTCATTACTGGAACACTCTGCTAGAAAGTTTACCCTGAGTACTTGGACCGGGAACAAT 695
QY AATTCAACTTCAGGG 1343
Db ACAAATTTCACTGG 711

RESULT 12
BY756516
LOCUS BY756516 RIKEN full-length enriched, blastocyst Mus musculus cDNA
DEFINITION BY756516 695 bp mRNA linear EST 17-DEC-2002
VERSION clone IL0020G08 3', mRNA sequence.
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H.,
Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C.,
Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A.,
Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S.,
Beisel,K.W., Blake,J.A., Brad,D., Brusic,V., Chothia,C., Corbani,
L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest,
A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A.,
Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J.,
Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M.,
King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons
P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki,
H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G.,
Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D.,
Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,
B.Z., Ringwald,M., Sandelin,A., Schneider,C., Sempile,C.A., Setou,
M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,
R.D., Tomita,M., Verardo,R., Wagner,L., Wahlstedt,C., Wang,Y.,
Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,
M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A.,
Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,
M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,
Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,
K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,
E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
22354683
12466851
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suenhiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.go.jp/
URL:http://genome.gsc.riken.go.jp/

```

```

Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,
S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K.,
Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y., Kondo,S., Konno,
H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K.,
Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N., Sano,H.,
Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Takeda,Y.,
Waki,K., Watanabe,A., Muramatsu,M. and Hayashizaki,Y. Direct
Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
FEATURES
source
Location/Qualifiers
1..695
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IL0020G08"
/dev_stage="blastocyst"
/clone_lib="RIKEN full-length enriched, blastocyst"
BASE COUNT 218 a 128 c 157 g 191 t 1 others
ORIGIN
Query Match 11.2%; Score 585.2; DB 14; Length 695;
Best Local Similarity 90.1%; Pred. No. 1.7e-109;
Matches 626; Conservative 0; Mismatches 69; Indels 0; Gaps 0;
QY 1809 TCATGTAGCTTTAAGCAGAGCTGGTCTCTTTTCAAGACTGCATGAATTTGTCTTTGA 1868
Db 1 TCATGTAGCTTTAAGCAGACTAGTGTCTATTTCAAGACTGCATGAATTTGTCTTTGA 60
QY 1869 GCACCTTCAAGTAGAGTACTAGTGAATATCTTTACCTTGTCTGTGTGTGTCGCCA 1928
Db 61 CGGCTTTCAAGTAGAGTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 120
QY 1929 GGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1988
Db 121 GGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 180
QY 1989 CCAAGATGTTAAGTCAGAGAAATATGTATGATAAGATATCATCATCATCATCATCAT 2048
Db 181 TCAAGATGTTAATTCAGGAGGAAATGTACGATAAGATATCATCATCATCATCATCAT 240
QY 2049 TGCATCTTTAATGGATCCCAATGAATTTCTTTACTGGTACTTCAGAGTATGAATTC 2108
Db 241 AGCATCTTTAATGGATCCCAATGAATTTCTTTACTGGTACTTCAGAGTATGAATTC 300
QY 2109 CGAGGCTTTAACAAGACCATATCTCAAAAGACAGGATTTTCATTAACAATATATAC 2168
Db 301 TGATGCTTTTAAACAAGACCATATCCAAAGACAGGATTTTCATTAACAAGTATATAC 360
QY 2169 ACTAATAGAAGAAATGCTTCAGGTCCTCATCTATATTTGGTGGTGGTGTATGACATGG 2228
Db 361 ATTATAGAAGAAATGCTTCAGGTCCTCATCTATATTTGGTGGTGGTGTATGACATGG 420
QY 2229 AGTGGAAATGTGACCAAGAGAGGTCAATGAGAGAAATCATTCATCTGCTTTGTCAT 2288
Db 421 AGTGGAAATGTGACCAAGAGAGGTCAATGAGAGAAATCATTCATCTGCTTTGTCAT 480

```


JOURNAL
MEDLINE
PUBMED

Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636

REFERENCE
AUTHORS

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)

20499374

JOURNAL
MEDLINE

11042159

REFERENCE
AUTHORS

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuina, T., Rashino, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashikawa, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

TITLE

20530913

JOURNAL
MEDLINE

11076861

REFERENCE
AUTHORS

Kawai, J., Shingawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamataka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Quackenbush, J., Schram, L. M., Staubli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bona, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seva, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohsaki, S. and Hayashizaki, Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)

TITLE

21085660

JOURNAL
MEDLINE

11217851

REFERENCE
AUTHORS

The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

TITLE

6 (bases 1 to 3058)

JOURNAL
MEDLINE

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp).

TITLE
JOURNAL

360

REFERENCE
AUTHORS

4934

REFERENCE
AUTHORS

420

COMMENT

URL: <http://genome.gsc.riken.go.jp/>, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL: <http://genome.gsc.riken.go.jp/>
URL: <http://fantom.gsc.riken.go.jp/>

FEATURES

Location/Qualifiers

1..3058

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="FANTOM_DB:6430402E23"

/db_xref="taxon:10090"

/clones="6430402E23"

/sex="male"

/tissue_type="olfactory brain"

/clone_lib="RIKEN full-length enriched mouse cDNA library"

/dev_stage="adult"

<1..691

/note="unnamed protein product: putative

ubiquitin protein ligase E3 component n-recognin 1

(MGDI|MF:1277977, GI:NM_009461, evidence: BLASTN, 100%, match=2560)

/codon_start=2

/protein_id="BAC37160.1"

/db_xref="GI:26347023"

/translation="GVAPPPELFANSARGESALCSYLSLTNLFLLQEWDTIRPL

LQWGDPAALSKQKQAVVYPRKNSLIEDPYSCLLNQASHFRCPKRSADDERK

HPVLCPGAILCSQICQVEVGCACVHACGAGVCIFLKTRECRVLLVEG

KARGCAPYLDVEYGETDPLKRGNPLHLRSRYRKLHLVQOHCIIIEIARSOETN

QMLFGFNWLL"

3032..3037

/note="putative"

3058

/note="putative"

BASE COUNT 870 a 562 c 632 g 994 t

ORIGIN

Query Match 11.1%; Score 577.4; DB 11; Length 3058;

Best Local Similarity 89.7%; Pred. No. 8.9e-108;

Matches .620; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

Qy 4515 TGGGTAACTCCGCTGAGGAATGCATACCAATTTCTGCAGAGGAGTACAGTGCCT 4574

1 TGGAGTAGCTCCGCTCGAGAACCTGTTGCCAATTTCTGCTGAGGAGATTCAGTGCCT 60

Qy 4575 CTGTAGCTATCTATCTTTTACCTACAAATTTGTTCTCTCCAGGAAATATTGGATAC 4634

61 CTGTAGCTATCTATCTTTTACCCACAAATTTGTTCTCTCTTTCCAGGAATATTGGATAC 120

Qy 4635 TGTAAAGCCCTTGCTCCAGAGCGGTGTGCAGATCTCGCTTACTTAAGCTGTTGAGCA 4694

121 CATAAGGCCCTTACTACAGAGGTGTGTGGAGATCTCGCTTACTCAAGCTTTTGAAGCA 180

Qy 4695 AAAAAACACCGCTGGTGCAGTACCCTAGAAAAAGAAATAGTTTGATAGAGCTTCTCTGATCA 4754

181 GAAAGTCTGTGGTGCAGTACCCCTAGAAAAAGAAATAGTTTGATAGAGCTTCTCTGATCA 240

Qy 4755 CTATAGCTTCCTCTCTGAATCAAGCTTCTCATTTTCAGGTGCCACGGTCTTCAGATGATGA 4814

241 CTACAGCTGTCTTCTAAATCAGGCTTCTCACATTTAGGTGTCCACGGTCTTCAGATGATGA 300

Qy 4815 GCGAAGCATCTGTCTCTCTGCTTCTGCTGGGCTATACATGTTCTCAGAACATTTG 4874

301 GCGAAGCATCTGTCTCTCTGCTTCTGCTGGGCTATACATGTTCTCAGAACATTTG 360

Qy 4875 CTGCCAGCAATTTGGAACGGGAGAGTGTGAGCTTGCAGCTTCTTTTCACGCACTTCTCAGT 4934

361 TTGGCAAGAAATAGTGAATGGGAGAGGTTGGAGCGTGGCGTTCATGCGCTTCATG 420

/clone_lib="NIH_BMAP_FD0"
/note="Organ: brain; Vector: pYX-Asc; Site_1: Ecor I;
Site_2: Not I; The library was constructed according to
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured mRNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with an
oligo-dT primer containing a Not I site. Double stranded
cDNA was size selected according to mRNA size fraction,
ligated with Ecor I adaptor, digested with Not I, and then
cloned directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
, is TGAGAGAGCC. This library was created for the
University of Iowa Mouse Brain Molecular Anatomy project

Search completed: September 27, 2003, 13:31:05
Job time : 6523.54 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 25, 2003, 14:45:45 ; Search time 104 Seconds
(without alignments)
2646.460 Million cell updates/sec

Title: US-09-724-126A-19

Perfect score: 9141

Sequence: 1 AMGNWADEAGTEREIS.....EIARSQTNMLFGFNWOLL 1734

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_19Jun03.*
1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9031	98.8	1738	24	ABP58330 Human cell growth,
2	8480	92.8	1757	20	AAW84351 Murine ubiquitin-p
3	8480	92.8	1757	22	AAB31162 Amino acid sequenc
4	4209	46.0	811	22	AAB93464 Human protein sequ
5	3359	36.7	1400	22	AAW78576 Human protein SEQ
6	3356	36.7	1400	22	AAW79560 Human protein SEQ
7	2612.5	28.6	1109	20	AAW02376 Polypeptide identi
8	2210.5	24.2	1829	22	ABB64216 Drosophila melanog
9	2168	23.7	424	23	ABB98104 Human ubiquitin re

10	1716	18.8	333	20	AAW84353 Partial human ubiq
11	1716	18.8	333	22	AAB31163 Amino acid sequenc
12	1071.5	11.7	452	23	ABB90168 Novel human diagno
13	877	9.6	487	22	ABG05917 Novel human diagno
14	790	8.6	258	22	ABG05881 Novel human diagno
15	606.5	6.6	1225	23	ABG93333 Herbicidally activ
16	449	4.9	389	22	ABG06664 Novel human diagno
17	414.5	4.5	203	23	ABB89137 Human polypeptide
18	408.5	4.5	951	24	ABJ26663 Human protein modi
19	362	4.0	1456	22	ABB58229 Drosophila melanog
20	360	3.9	153	22	AAH89524 Human immune/haema
21	311	3.4	783	22	ABE58228 Drosophila melanog
22	251	2.7	250	22	AAW25572 Human protein sequ
23	250	2.7	247	21	ABP38334 Human secreted pro
24	240	2.6	81	24	ABP54255 Human lung specifi
25	208	2.3	2230	24	ABU07445 Protein differenti
26	197.5	2.2	2096	21	ABH41592 Human ORFX ORF1356
27	196	2.1	953	22	ABU53071 Intracellular traf
28	189.5	2.1	2048	22	ABM40027 Human polypeptide
29	186	2.0	951	22	ABU53070 Intracellular traf
30	184	2.0	1411	17	AAW02258 Nucleolar/endosoma
31	182	2.0	2442	21	AAW77575 Human cytoskeletal
32	180	2.0	996	22	AAU33755 Staphylococcus aur
33	180	2.0	1009	22	AU36548 Staphylococcus aur
34	179.5	2.0	2856	24	ABU57575 Mouse lrbA protein
35	178.5	2.0	961	22	ABU53077 Intracellular traf
36	177.5	1.9	2816	22	AAU68572 Human novel cytoki
37	177	1.9	1373	24	ABU11772 Human MDTF polypep
38	176	1.9	2013	22	ABE62322 Drosophila melanog
39	176	1.9	2017	22	ABG06301 Novel human diagno
40	176	1.9	2888	22	AAW40883 Human polypeptide
41	175.5	1.9	1427	12	AAW10534 Human 160KB mediat
42	175.5	1.9	2633	22	ABG06505 Novel human diagno
43	175.5	1.9	2663	22	AAW39097 Human polypeptide
44	175	1.9	2383	23	ABG65631 Human breast speci
45	174.5	1.9	1392	20	AAW06999 Restin protein seq

ALIGNMENTS

RESULT 1

ABP58330
ID ABP58330 standard; Protein; 1738 AA.

XX AC ABP58330;

XX DT 07-APR-2003 (first entry)

XX DE Human cell growth, differentiation and death protein CGDD-1.

XX DE CGDD-1; cell growth, cell differentiation; cell death; human; cytostatic; antiarteriosclerotic; hepatotropic; antiinflammatory; antipsoriatic; antianaemic; ophthalmological; auditory; anticonvulsant; cerebroprotective; nootropic; neuroprotective; antiparkinsonian; neuroleptic; tranquilizer; immunosuppressive; anti-HIV; antiallergic; antitachymatic; antithyroid; antidiabetic; dermatological; nephrotropic; antirheumatic; antiarthritic; antiulcer; vulnerary; virucide; antibacterial; fungicide; antiparasitic; protozoacide; antihelminthic; antinfertility; gynaecological; ubiquitin protein ligase; enzyme; gene therapy.

XX OS Homo sapiens.

XX PN WO200297032-A2.

XX PD 05-DEC-2002.

XX PF 05-APR-2002; 2002WO-US111152.

XX PR 06-APR-2001; 2001US-282110P.

XX PR 11-APR-2001; 2001US-283294P.

XX PR 26-APR-2001; 2001US-286820P.

PR 27-APR-2001: 2001US-287228P.
PR 16-MAY-2001: 2001US-291562P.
PR 18-MAY-2001: 2001US-291846P.
PR 25-MAY-2001: 2001US-293727P.
PR 01-JUN-2001: 2001US-295263P.
PR 01-JUN-2001: 2001US-295340P.
PR 15-JAN-2002: 2002US-349705P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Azimzai Y, Au-Young JK, Batra S, Baughn MR, Becha SD, Borowsky KJ,
PI Burford N, Ding L, Elliott VS, Emerling BM, Gandhi AR, Gietzen MJ,
PI Griffin JA, Hafalia AJA, Honchell CD, Lal PG, Lee SY, Lu DM,
PI Arvizu CS, Ramkumar J, Reddy R, Sanjanwala MM, Tang YT, Walla NK,
PI Wang YE, Warren BA, Xu Y, Yang J, Yao MG, Yue H, Zebardjian Y;
XX
DR WPI; 2003-140453/13.
DR N-PSDB; AB224689.
XX
PT Novel human proteins associated with cell growth, differentiation and
PT death, useful for treating, diagnosing or preventing cancer,
PT developmental, neurological, reproductive or autoimmune/inflammatory
PT disorders -
XX
XX Claim 1; Page 183-187; 238pp; English.
PS
XX The present sequence is the protein sequence of human CGDD-1, a
CC novel protein associated with cell growth, differentiation and
CC death. The sequence is predicted from Incyte clone 1351608CB1,
CC which was isolated from a paraneoplastic tumour tissue CDNA
CC library. Structural features establish the protein as being
CC associated with cell growth, differentiation and death, with
CC further evidence suggesting it to be a ubiquitin protein ligase.
CC The invention is based on novel human CGDD-1 to -21 proteins (see
CC ABP58330-50), the polynucleotides encoding them (see AB224689-709),
CC and to the use of these for the diagnosis, treatment or prevention
CC of cell proliferative disorders including cancer, developmental
CC disorders, neurological disorders, autoimmune disorders,
CC reproductive disorders, and disorders of the placenta, and in the
CC assessment of the effects of exogenous compounds on the activity
CC and expression of proteins and nucleic acids associated with cell
CC growth, differentiation and death.
XX
SQ Sequence 1738 AA;
Query Match 98.8%; Score 9031; DB 24; Length 1738;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1717; Conservative 0; Mismatches 1; Indels 20; Gaps 1;
QY 17 MEISALPQTPORLASWMDQVDFTAFLLHLLAQLVPEIYFAEMDPDLEKQESVQMSIF 76
DB 1 MEISALPQTPORLASWMDQVDFTAFLLHLLAQLVPEIYFAEMDPDLEKQESVQMSIF 60
QY 77 TPLEWYLFGEDDPDLCKLHSGAFOLCGRVFKSGGTTYSRCDAIDPTCVLCMDCFQDS 136
DB 61 TPLEWYLFGEDDPDLCKLHSGAFOLCGRVFKSGGTTYSRCDAIDPTCVLCMDCFQDS 120
QY 137 VKNHRYKWHSTGGGFCDCGDTFAWKTPFCVNHPEPGRAGTIKENSRCPLNEEIVQAR 196
DB 121 VKNHRYKWHSTGGGFCDCGDTFAWKTPFCVNHPEPGRAGTIKENSRCPLNEEIVQAR 180
QY 197 KIPPSVIKYVEMTIWEEKEKLPPELQIREKNERYCYVLFNDRHHSDYHVIYSIQRALDC 256
DB 181 KIPPSVIKYVEMTIWEEKEKLPPELQIREKNERYCYVLFNDRHHSDYHVIYSIQRALDC 240
QY 257 ELAEALHTTAIDKEGRRAVKAGAYAAQCEAKEDIKSHSNVSQHPHLEVLHSEIMAHQ 316
DB 241 ELAEALHTTAIDKEGRRAVKAGAYAAQCEAKEDIKSHSNVSQHPHLEVLHSEIMAHQ 300
QY 317 KPALRGSMWNKIMSYSSDFRQIFCOACLRPEPDSNPCLISRLMLWDKLYKGARKILH 376
DB 301 KPALRGSMWNKIMSYSSDFRQIFCOACLRPEPDSNPCLISRLMLWDKLYKGARKILH 360

QY 377 ELIFSSFFMEMEYKKLFAMEFVKYKOLQKEYISDDHDSRISITALSVMQFTVPTLARHL 436
DB 361 ELIFSSFFMEMEYKKLFAMEFVKYKOLQKEYISDDHDSRISITALSVMQFTVPTLARHL 420
QY 437 IEEQNVISVITETILLEVLPEYLDNRNKNFQGYSDQKLGRIYAVICDLKYILISKPTIWT 496
DB 421 IEEQNVISVITETILLEVLPEYLDNRNKNFQGYSDQKLGRIYAVICDLKYILISKPTIWT 480
QY 497 ERLRMOPLGFRSFLKILTQMGMEETRRQVGHIEVDPDWEAAIAIOMQKNILLMPQE 556
DB 481 ERLRMOPLGFRSFLKILTQMGMEETRRQVGHIEVDPDWEAAIAIOMQKNILLMPQE 540
QY 557 WCACDELLAVAYKECHKAVMRCSTSFSSSKTVQSCGSHLETKSVRSVSDLSIHLPL 616
DB 541 WCACDELLAVAYKECHKAVMRCSTSFSSSKTVQSCGSHLETKSVRSVSDLSIHLPL 600
QY 617 SRTLGLHVRSLRGAVSRLHEFVSFEDFQVEVLVEYPLRCLVLVAQVVAEMWRNGLSL 676
DB 601 SRTLGLHVRSLRGAVSRLHEFVSFEDFQVEVLVEYPLRCLVLVAQVVAEMWRNGLSL 660
QY 677 ISOVFFYQDYKCREMYDKDIIMLQIGASLMDPNKFLLLVQLRYELAEAFNKTISTKDDQ 736
DB 661 ISOVFFYQDYKCREMYDKDIIMLQIGASLMDPNKFLLLVQLRYELAEAFNKTISTKDDQ 720
QY 737 LIQOYNTLIBEMLQVLIYIGERYVPGVGNVTKEEVTMREIILHLLCTEPMHPSATAKNLP 796
DB 721 LIQOYNTLIBEMLQVLIYIGERYVPGVGNVTKEEVTMREIILHLLCTEPMHPSATAKNLP 780
QY 797 ENENNETGLENVINKVATFKKPGVSGHGVYELKDESLKDFNMNFYHYSKTOHSAEHMOK 856
DB 781 ENENNETGLENVINKVATFKKPGVSGHGVYELKDESLKDFNMNFYHYSKTOHSAEHMOK 840
QY 857 KRRKQENKDEALPPPPPPPCPAFSKVINLNCDDIMYILRTVFERAIDTDSNLTGML 916
DB 841 KRRKQENKDEALPPPPPPPCPAFSKVINLNCDDIMYILRTVFERAIDTDSNLTGML 900
QY 917 QMAFHIALGLLEKQOLQKAPAEVTFDFYHKASRLGSSAMNTOMLLEKLGIPQLBQ 976
DB 901 QMAFHIALGLLEKQOLQKAPAEVTFDFYHKASRLGSSAMNTOMLLEKLGIPQLBQ 960
QY 977 KDMITWILQMFDTVKRLREKSCLLVATTSGSESINKDEITHDKAEKRAEAAARLRQ 1036
DB 961 KDMITWILQMFDTVKRLREKSCLLVATTSGSESINKDEITHDKAEKRAEAAARLRQ 1020
QY 1037 KINQAQMSALQKNFTETHKLMYDNTSEMPGKEDSIMEESTPAVSYSRIALGPKRGPSVT 1096
DB 1021 KINQAQMSALQKNFTETHKLMYDNTSEMPGKEDSIMEESTPAVSYSRIALGPKRGPSVT 1080
QY 1097 EKEVLTICLQCEOEVEKIENNAMVLSACVQKSTALTQHRGKPIELSGEALDPLFMDPLA 1156
DB 1081 EKEVLTICLQCEOEVEKIENNAMVLSACVQKSTALTQHRGKPIELSGEALDPLFMDPLA 1140
QY 1157 YGVTGSCGHVMHACVQKFEAVOLSSQQRHVDLFDLESGEYLCPLCKSLCNTVPIPI 1216
DB 1141 YGVTGSCGHVMHACVQKFEAVOLSSQQRHVDLFDLESGEYLCPLCKSLCNTVPIPI 1200
QY 1217 PLOPKINSENADALAQLLTARWIQTVARISGYNIRHAKGENPIPIFFNOGMDSTLE 1276
DB 1201 PLOPKINSENADALAQLLTARWIQTVARISGYNIRHAKGENPIPIFFNOGMDSTLE 1260
QY 1277 FHSILSGVSESSIKYSNIKEMVILFATTIYRIGLKVPPDERDPRVPMLTWSTCAFTQA 1336
DB 1261 FHSILSGVSESSIKYSNIKEMVILFATTIYRIGLKVPPDERDPRVPMLTWSTCAFTQA 1320
QY 1337 IENLLGDEGKPLFGALONROHNGKALMQFAVORITCPQVLIQKHLVRLSVLPNPKS 1396
DB 1321 IENLLGDEGKPLFGALONROHNGKALMQFAVORITCPQVLIQKHLVRLSVLPNPKS 1380
QY 1397 EDTPCLLSIDLPHVLVGAVALFAPSLYWDPPDLPQSSVSSSYNHLFLPHLTMAHMLQIL 1456
DB 1381 EDTPCLLSIDLPHVLVGAVALFAPSLYWDPPDLPQSSVSSSYNHLFLPHLTMAHMLQIL 1440
QY 1457 LTVDTGLPLAQVQEDSEAHSSAFFAEISQYTSIGSICDDIPGWLYWLSLKNGITPYLR 1516

```

Db 1441 LTVDGLPLAQVQDSSEAHSSAFELISQYTSIGCDIPGWLWLSKNGITPYLRC 1500
QY 1517 AALFFHYLLGVTTPPELHTNSAEGEYSALCSYLSLPTNLFLLFQYWDVTRPQLLQRRCAD 1576
Db 1501 AALFFHYLLGVTTPPELHTNSAEGEYSALCSYLSLPTNLFLLFQYWDVTRPQLLQRRCAD 1560
QY 1577 PALLNCLKQKNTVVRYPRKRNLSIELPDDYSCLLNOASHFRCPRSADDERKHPVLCFLFCG 1636
Db 1561 PALLNCLKQKNTVVRYPRKRNLSIELPDDYSCLLNOASHFRCPRSADDERKHPVLCFLFCG 1620
QY 1637 AILCSNCCOEIVNGEEVGACIFHALHC-----KARGCAYPAPY 1676
Db 1621 AILCSNCCOEIVNGEEVGACIFHALHCAGVCIFLIRECRVVLVEGKARGCAYPAPY 1680
QY 1677 LDEYGETDPGLKRGNPLHLRSERYKRLHLVMOQHCHIEEIRASQETNOMLFQFNWOLL 1734
Db 1681 LDEYGETDPGLKRGNPLHLRSERYKRLHLVMOQHCHIEEIRASQETNOMLFQFNWOLL 1738

RESULT 2
AAW84351
ID AAW84351 standard; Protein; 1757 AA.
AC AAW84351;
XX
DT 25-MAR-1999 (first entry)
DE Murine ubiquitin-protein ligase Ubr1.
XX Ubiquitin-protein ligase; Ubr1; mouse; ubiquitinylation; degradation;
KW N-end rule pathway; stress-related muscle wasting; inhibitor; screen.
XX
OS Mus sp.
XX
PN US5861312-A.
XX
PD 19-JAN-1999.
XX
PF 02-DEC-1997; 97US-0982956.
XX
PR 02-DEC-1997; 97US-0982956.
XX
PA (CALY ) CALIFORNIA INST OF TECHNOLOGY.
XX
PI Kwon YT, Varshavsky A;
XX
DR WPI; 1999-130395/11.
DR N-PSDB; AAV99308.
XX
PT Mouse and human Ubr1 cDNA - useful for producing recombinant Ubr1
polypeptides
XX
PS Disclosure; Columns 15-28; 18pp; English.
XX
CC The present sequence represents a ubiquitin-protein ligase called Ubr1.
CC The Ubr1 enzymes are involved in protein ubiquitinylation and
CC ultimate degradation through the N-end rule pathway and have been
CC linked to stress-related muscle wasting. Recombinant Ubr1 polypeptides
CC can be used to screen for inhibitors of muscle wasting when this is
CC associated with the N-end rule pathway.
XX
SQ Sequence 1757 AA;

Query Match 92.8%; Score 8480; DB 20; Length 1757;
Best Local Similarity 91.0%; Pred. No. 0;
Matches 1598; Conservative 64; Mismatches 67; Indels 28; Gaps 4;

QY 6 MADEAGGTERMEISAEPLQTPORLASWDDQVDYTFALHHLAQLVPEIFAEWMDPDLE 65
Db 1 MADEMDGAEKMDVSPPEPLAPQRPASWDDQVDYTFALHHLAQLVPEIFAEWMDPDLE 60
QY 66 KOESVQMSITPLEWYLFGEPPDICLEKLHSGAFOLCGRVKSGSETTYSRCDAIDPT 125

```

```

Db 61 KOESVQMSITPLEWYLFGEPPDICLEKLHSGAFOLCGRVKSGSETTYSRCDAIDPT 120
QY 126 CVLCMDCFOQSVHKNHRYKMHSTGTGGFCDCGDTFAMKTPFCVNHHEPRAGTITKENSRC 185
Db 121 CVLCMDCFOQSVHKNHRYKMHSTGTGGFCDCGDTFAMKTPFCVNHHEPRAGTITKESLHC 180
QY 186 PLNEEVIVQARKTIPSPVIVKVVEMTWIWEKEKELPELQIREKNERYCVLENDEHHSYDH 245
Db 181 PLNEEVIAQARRIFPSPVIVKVVEMTWIWEKEKELPELQIREKNERYCVLENDEHHSYDH 240
QY 246 VIYSLOALDCELAEAQLHTTADKEGRRVAVKAGAYAACQAEKEDIKSHSENVSOHPLHV 305
Db 241 VIYSLOALDCELAEAQLHTTADKEGRRVAVKAGAYATCQAEKEDIKSHSENVSOHPLHV 300
QY 306 EVLHSEITMAHOKFALRLGSMWNKIMSYSSQFQFCOACLRPEEDSENPCILSRMLMWA 365
Db 301 EVLHVVMAHOKFALRLGSMWNKIMSYSSQFQFCOACLRPEEDSENPCILSRMLMWA 360
QY 366 KLYKGARKILHELIFSSFFMEMEYKLFAMEFVYKQLOKEYISDDHDSISITALSVQ 425
Db 361 KLYKGARKILHELIFSSFFMEMEYKLFAMEFVYKQLOKEYISDDHDSISITALSVQ 420
QY 426 MFTVPTLARHLIEEQNVISVITETLLEVLPEYLDNRNKNFQGYSDQKLGRYAVICDLK 485
Db 421 MFTVPTLARHLIEEQNVISVITETLLEVLPEYLDNRNKNFQGYSDQKLGRYAVICDLK 480
QY 486 YLISKPTIWTERTLRMOFLEGFRSFLKILTCMOGMEIRQVGOHIEVDPDWEAAIATQM 545
Db 481 YLISKPTIWTERTLRMOFLEGFRSFLKILTCMOGMEIRQVGOHIEVDPDWEAAIATQM 540
QY 546 QLKNIILLMFQEWACDEELLLVAYKECHKAVMRCTSFISSTKTVVSCGHSLETYSRV 605
Db 541 QLKNIILLMFQEWACDEELLLVAYKECHKAVMRCTSFISSTKTVVSCGHSLETYSRV 600
QY 606 SEDLVSIHLPLSRTLAGLHVRLSRLGAVSRHLEFVSFEDQVEVLYEPLRCLVLVAQVV 665
Db 601 SEDLVSIHLPLSRTLAGLHVRLSRLGAVSRHLEFVSFEDQVEVLYEPLRCLVLVAQVV 660
QY 666 AEMWRRNGLSLSQVFFYQDVCKREEMYDKDIIMLOIGASIMDPNKFLLVLQRYELAE 725
Db 661 AEMWRRNGLSLSQVFFYQDVCKREEMYDKDIIMLOIGASIMDPNKFLLVLQRYELAE 720
QY 726 FNKTISTKDQDLIKQYNTLIEEMLOVLIYIVGERYPVGVGNVTEEMTIRIHLCTEP 785
Db 721 FNKTISTKDQDLIKQYNTLIEEMLOVLIYIVGERYPVGVGNVTEEMTIRIHLCTEP 780
QY 786 MPHSAITAKNLPENENNETGLENVINKVATFKPGVSGHGYVELKDESJKDFNMFFYHYSK 845
Db 781 MPHSAITAKNLPENENNETGLENVINKVATFKPGVSGHGYVELKDESJKDFNMFFYHYSK 840
QY 846 TQHSKAEHMOKKRRKQENKDEALPPPPPECFPAFSKVINLLNCDIMWYILRTVFERAID 905
Db 841 TQHSKAEHMOKKRRKQENKDEALPPPPPECFPAFSKVINLLNCDIMWYILRTVFERAID 900
QY 906 TQSNLWTEGMLQAFHILALGLEEKQOLKAPEEVTFDFYHKASRLGSSAM--NIQM 962
Db 901 TQSNLWTEGMLQAFHILALGLEEKQOLKAPEEVTFDFYHKASRLGSSAMNIQM 960
QY 963 LLEKLKGIPOLEGQKDMITWILQMFDTVKRLREKSLIVATTSGESIKNDEITHDKKA 1022
Db 961 LLEKLKGIPOLEGQKDMITWILQMFDTVKRLREKSLIVATTSGESIKNDEITHDKKA 1020
QY 1023 ERKRKAELARHROKIMQASALQKNPIETHKLMYDNTSEMPGKEDSIMEESTPAYSVDY 1082
Db 1021 ERKRKAELARHROKIMQASALQKNPIETHKLMYDNTSEMPGKEDSIMEESTPAYSVDY 1080
QY 1083 SRTALGPKRGPSVTEKEVLTCILCOEVEQVKIENNAVLSACVOKSTALTQHRKPKLELS 1142
Db 1081 SRTALGPKRGPSVTEKEVLTCILCOEVEQVKIENNAVLSACVOKSTALTQHRKPKVDHL 1140
QY 1143 GEALDPLFMDPDLAYGTYTGTSCGRVMAHVCWQKYFEAVQLSSQQRHVDLFDLESGEYLC 1202

```

Db 1141 GETLDPLFMDPDLAHGTYTSCGHVMHVCWQKYFEAVQLSSQRIHVLDLFDLESGEYLC 1200
 Qy 1203 PLCKSLCNVTIPIIPIQOKINSENADALQALLTLARWITQVTLARISGVNIRAKGNP- 1261
 Db 1201 PLCKSLCNVTIPIIPIQOKINSENADALQALLTLARWITQVTLARISGVNIRAKGNP 1260
 Qy 1262 IPIFFNOGMDSTLEPHSILSFQVSSIKYSIKEMVILFATYIRIGLKVPPDERDP 1321
 Db 1261 VPVLFNOGMDSTLEPHSILSFQVSSIKYSIKEMVILFATYIRIGLKVPPDELDP 1320
 Qy 1322 VPMLTWSTCAFTIQAIENTLLGDEKPLFGALQNRQINGKALMQFAVAQAITCPQVLIQ 1381
 Db 1321 VPMWSTCAFTIQAIENTLLGDEKPLFGALQNRQINGKALMQFAVAQAITCPQVLIH 1380
 Qy 1382 HLVRLLSVLPNTKSEDTPCCLSIDLFLHVLGAVLAPPSLYWDPDVQLPSSVSSYNHL 1441
 Db 1381 HLRLLSVLPNTKSEDTPCCLSIDLFLHVLGAVLAPPSLYWDPDVQLPSSVSSYNHL 1440
 Qy 1442 YLPHLITMAHMLQILLTVDI---GLPLAQVQEDSEAHSAFFFAISQYTSIGICDI 1497
 Db 1441 YLPHLITMAHMLQILLTDTLSPGPPAEGEEDSEARCAFAFFVEVSQHTDGLTCCGA 1500
 Qy 1498 PGYLMWSLKNGITPYLRCAALFFHYLLGVTPPEELHTNSAEGYSALCSYLSLPTNLF 1557
 Db 1501 PGYLMWSLKNGITPYLRCAALFFHYLLGVTPPEELHTNSAEGYSALCSYLSLPTNLF 1560
 Qy 1558 LFOEYWDTRVPLQRRCADPALLCNLCLOKNTVYPRKRNSLIELPDDYSCLLNQASHFR 1617
 Db 1561 LFOEYWDTRVPLQRRCADPALLCNLCLOKNTVYPRKRNSLIELPDDYSCLLNQASHFR 1620
 Qy 1618 CPRSADDERKHPVLCFCGAILCSQICQOEIYNGVEGVGACIFALHC----- 1665
 Db 1621 CPRSADDERKHPVLCFCGAILCSQICQOEIYNGVEGVGACIFALHC----- 1680
 Qy 1666 -----KARGCAYPAPYLDEYGETDPGLKRGNPLHLRSRYRKHLVWQHCIIIEIA 1717
 Db 1681 CRVVLVEGKARGCAYPAPYLDEYGETDPGLKRGNPLHLRSRYRKHLVWQHCIIIEIA 1740
 Qy 1718 RSQETNOMLFGFNWQLL 1734
 Db 1741 RSQETNOMLFGFNWQLL 1757

RESULT 3

AAB31162
 ID AAB31162 standard; Protein: 1757 AA.

AC AAB31162;

XX 02-APR-2001 (first entry)

DT Amino acid sequence of Mouse Ubr1 protein.

DE Ubr1; E3-type protein; ubiquitin system; ubiquitin-protein ligase;
 KW N-end rule pathway; intracellular pathogen; Lysteria monocytogenes;
 KW Yersinia enterocolitica; muscle wasting; infection.

XX Mus sp.

XX US6159732-A.

XX 12-DEC-2000.

PD 11-JAN-1999; 99US-0228317.

PF 02-DEC-1997; 97US-0982956.

PR (CALY) CALIFORNIA INST OF TECHNOLOGY.

PA Kwon YT, Varshavsky A;

PI WPI; 2001-090278/10.

XX N-PSDB; AAC86933.

DR

XX Inhibiting the N-end rule pathway in mammalian cells for treating
 PT infections and various diseases associated with muscle tissue wasting,
 PT by inhibiting the expression of Ubr1 gene -
 XX
 PS Example; Column 15-28; 18pp; English.

XX The present sequence represents a murine Ubr1 enzyme. Ubr1 is an E3-type
 CC protein of the ubiquitin system. Specifically, it is a ubiquitin-protein
 CC ligase. The enzyme is specific for destabilising residues exposed at
 CC the N-terminus of protein substrates. Inhibition of the expression of
 CC Ubr1 gene in a cell results in inhibition of the N-end rule pathway.
 CC The method is used for treatment of mammalian cells infected with an
 CC intracellular pathogen, e.g. Lysteria monocytogenes or Yersinia
 CC enterocolitica. Inhibition of N-end rule pathway is also useful for
 CC treating various diseases associated with wasting of muscle tissue and
 CC infections.

XX SQ Sequence 1757 AA;

Query Match 92.8%; Score 8480; DB 22; Length 1757;
 Best Local Similarity 91.0%; Pred. No. 0;
 Matches 1598; Conservative 64; Mismatches 67; Indels 28; Gaps 4;

Qy 6 MADEAGCTERMEISAEELPQTPORLASWMDQVDYTFATLHHLAQLVPEIYFAEMDPDLE 55
 Db 1 MADEEMDGAERMDVSPEPLAPQRPASWMDQVDYTFATLHHLAQLVPEIYFAEMDPDLE 60
 Qy 66 KOESVQMSIITPLEWYLFEGDDPDICLEKLGAFQVLCGRVFKSGGETTSCRDCAIDPT 125
 Db 61 KOESVQMSIITPLEWYLFEGDDPDICLEKLGAFQVLCGRVFKSGGETTSCRDCAIDPT 120
 Qy 126 CVLQMDQFQDSVHKNHRYKMHMTSTGGGFCDCGDTAEWTKPFCVNHPEFRAGTIKENSRC 185
 Db 121 CVLQMDQFQDSVHKNHRYKMHMTSTGGGFCDCGDTAEWTKPFCVNHPEFRAGTIKENSRC 180
 Qy 186 PLNEEVIVQARKIPSPVYKIVVEMTIWEEELPELOIREKNERYCYVLFNDEHSHSDH 245
 Db 181 PLNEEVIVQARKIPSPVYKIVVEMTIWEEELPELOIREKNERYCYVLFNDEHSHSDH 240
 Qy 246 VIYSIQRALDCELAELAAQLHTTAIDKEGRRAVKAGAYACQAEKEDIKSHSENVSOHPH 305
 Db 241 VIYSIQRALDCELAELAAQLHTTAIDKEGRRAVKAGAYATCQAEKEDIKSHSENVSOHPH 300
 Qy 306 EVLHSEIMAHQKFAIRLGSMWNKIMSYSDFRQIFCQACLRPEEPSENPCILSRMLMDA 365
 Db 301 EVLHSEIMAHQKFAIRLGSMWNKIMSYSDFRQIFCQACLRPEEPSENPCILSRMLMDA 360
 Qy 366 KLYKGARKILHELIFSSFFMEMEYKKLFAMEFVYKYLQKEYISDDHRSISITALSVO 425
 Db 361 KLYKGARKILHELIFSSFFMEMEYKKLFAMEFVYKYLQKEYISDDHRSISITALSVO 420
 Qy 426 MFTVPTLARHLIEONVISVITETLLEVLPEYLDNRNKNFNFGYSQDKLGRVYAVICDLK 485
 Db 421 MFTVPTLARHLIEONVISVITETLLEVLPEYLDNRNKNFNFGYSQDKLGRVYAVICDLK 480
 Qy 486 YILISKPTIWTLRMQFLEGFRSLKILTCMQGMEIIRROVQGHIEVDPDWEAAIAIQM 545
 Db 481 YILISKPTIWTLRMQFLEGFRSLKILTCMQGMEIIRROVQGHIEVDPDWEAAIAIQM 540
 Qy 546 QLKNIILLFQEWACDEBLLLVAYKECHKAVNRCSTSFISSTKTVVQSGCHSLETKSYRV 605
 Db 541 QLKNIILLFQEWACDEBLLLVAYKECHKAVNRCSTSFISSTKTVVQSGCHSLETKSYRV 600
 Qy 606 SEDLVS IHLPLSRTLAGLHVRSLRGAVSRLEHFVSEDFQVEVLVEYPLRCLVLVAQVY 665
 Db 601 SEDLVS IHLPLSRTLAGLHVRSLRGAVSRLEHFVSEDFQVEVLVEYPLRCLVLVAQVY 660
 Qy 666 AEMWRRNGLSLTSQVFFYQDVKCREMYDKDILMLQIGASLMDPNKFLLLVLRQYELARA 725
 Db 661 AEMWRRNGLSLTSQVFFYQDVKCREMYDKDILMLQIGASLMDPNKFLLLVLRQYELARA 720
 Qy 726 FNKTIKTDQDLIKOYNLTLEMLQVLVIYIGERVYVPGVGNVTKSEVTMRETIHLLCIEP 785

```
Db 721 FNKTIKTDQDLIKQYNTLIEMLQVLIYIGERYVPGVGNVTRREVMREITHLLCIEP 780
QY 786 MPHSAIAKNLPENNENNETGLENVINKVATFKPGVSGHGVYELKDESLKDFNMYYHYSK 845
Db 781 MPHSAIARNLPENNENNETGLENVINKVATFKPGVSGHGVYELKDESLKDFNMYYHYSK 840
QY 846 TOHSAKAEHMOKRRKQENKDEALPPPPPEPCPAFASKVINLLNCDDIMMYILRTVFERAID 905
Db 841 TOHSAKAEHMOKRRKQENKDEALPPPPPEPCPAFASKVINLLNCDDIMMYILRTVFERAID 900
QY 906 TDSNLWTEGMLQMAFHIALGLLEKQLOKAPAEVEVTFDYHKASRLGSSAM---NIQM 962
Db 901 TESNLWTEGMLQMAFHIALGLLEKQLOKAPAEVEVTFDYHKASRLGSSAMNAQNIQM 960
QY 963 LLEKLGIPQLEGQKDMITWILQMFDTVKRUREKSLIVATTSGESIKNDEITHDKEKA 1022
Db 961 LLEKLGIPQLEGQKDMITWILQMFDTVKRUREKSLIVATTSGESIKNDEITHDKEKA 1020
QY 1023 ERKEKAEARLHRQKIMAQNSALQKNFTIETHKLMYDNTSEMPGKREDSIMEESTPAVSDY 1082
Db 1021 ERKKAARLHRQKIMAQNSALQKNFTIETHKLMYDNTSEYTGKREDSIMEESTSAVSEA 1080
QY 1083 SRIALGPKRGSPVTEKEVLTICLQEOEVEKIENNAMVLSACVOKSTALTOHRGKPIELS 1142
Db 1081 SRIALGPKRGSPVTEKEVLTICLQEOEVEKIENNAMVLSACVOKSTALTOHRGKPDHL 1140
QY 1143 GEALDPLFMDPDLAYGTGSCGHVMHMAVCWQKYFEAVQLSSQQRHIVDFLDESSEYLC 1202
Db 1141 GETLDPLFMDPDLARGTYTSCGHVMHMAVCWQKYFEAVQLSSQQRHIVDFLDESSEYLC 1200
QY 1203 PLKSLCNTVPIPIPLQPKINSNADALQALLLARIQVILARISGINYRHAKEGPN- 1261
Db 1201 PLKSLCNTVPIPIPLQPKINSNAEALQALLLARIQVILARISGINYRHAKEGAPA 1260
QY 1262 IPIPFNMGMDSTLEFHSILSFGVESSTIKYSNISKEMVILPATIYIRIGLVPPDERDP 1321
Db 1261 VPVLFNMGMDSTEFHSILSFGVQSSVKYSNISKEMVILPATIYIRIGLVPPDELPDR 1320
QY 1322 VPMLTWTCAFTIQAIEENLGDGKPLFGALQNRQHNKALMOFAVAQRITCPOVLIOK 1381
Db 1321 VPMMWTCAFTIQAIEENLGDGKPLFGALQNRQHSGLKALMOFAVAQRATCPOVLIOK 1380
QY 1382 HLVRLLSVLPNKSEDTPCLLSIDLHVLVGAVALPSPSLVWDVDPVLOPSSVSSYNHL 1441
Db 1381 HLARLLSVLPNLQSENTPGLLSVDLFHLVGVAVLAPSLXWDDTVLOPSPSSYNHL 1440
QY 1442 YLFHLITMAHMLQILLTVDOT---GLPLAQVQEDSEEAHSAFSAEISQYTSIGCDI 1497
Db 1441 YLFHLITMAHMLQILLTVDOTDLSPLGPPPLAEGEEDSEEARCASFVEYSQHTDGLTGCGA 1500
QY 1498 PGWYLMVSLKNGITPYLRCAALFFHYLLGVTPPELHTNSAEGYSALCSYLSLPTNLF 1557
Db 1501 PGWYLMVSLRNGITPYLRCAALLFHYLLGVAPPELFANSRGEFSALCSYLSLPTNLF 1560
QY 1558 LFOEYWDTVRPLQRRCADPALLNCLQKNTVVRYPKRNRSLIELPDDYSCLLNOASHFR 1617
Db 1561 LFOEYWDTVRPLQRRCADPALLNCLQKNTVVRYPKRNRSLIELPDDYSCLLNOASHFR 1620
QY 1618 CPRSADDERKHPVLCFGAILCSQNTCCQBIIVGEEVGACIFIALHC----- 1665
Db 1621 CPRSADDERKHPVLCFGAILCSQNTCCQBIIVGEEVGACIFIALHC----- 1680
QY 1666 -----KARGCAYPAPYLDXEYGETDPLGRGNPLHLRSERYRKLHLVWQOHCIIETIA 1717
Db 1681 CRVVLVEGKARGCAYPAPYLDXEYGETDPLGRGNPLHLRSERYRKLHLVWQOHCIIETIA 1740
QY 1718 RSQETNQLFGFNWQLL 1734
Db 1741 RSQETNQLFGFNWQLL 1757
```

AAB93464

ID AAB93464 standard; Protein; 811 AA.

XX AAB93464;

XX 26-JUN-2001 (first entry)

XX Human protein sequence SEQ ID NO:12732.

DE Human; primer; detection; diagnosis; antisense therapy; gene therapy.

KW Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isoqai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX

PR Primer sets for synthesizing polynucleotides, particularly the 5602

PT full-length cDNAs defined in the specification, and for the detection

PT and/or diagnosis of the abnormality of the proteins encoded by the

PT full-length cDNAs -

XX

PS Claim 8; SEQ ID 12732; 2537pp + CD ROM; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

SQ Sequence 811 AA;

Query Match 46.0%; Score 4209; DB 22; Length 811;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 811; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 209 MTIWEERKELPPELOIREKNERYCVLFNDEHSHYDHVYISLORALDCELAEOQLHTAI 268

Db 1 MTIWEERKELPPELOIREKNERYCVLFNDEHSHYDHVYISLORALDCELAEOQLHTAI 60

QY 269 DKGRRAVAGAYAAQCAKEADIKSHSNVYSHQPLHVEVLHSEIMAHQKFAALGLGWMNK 328

```
Db 61 DREGRAVAGAAQCEAKEDIKSHSENVSHPLHVEVLHSEIMAHQFALRGSMWNK 120
QY 329 IMSYSSDFRQICQACLRPEPSENCLISRLMLDAKLYKGARKILHELIFSSFFMEME 388
Db 121 IMSYSSDFRQICQACLRPEPSENCLISRLMLDAKLYKGARKILHELIFSSFFMEME 180
QY 389 YKKLFAMEFVKYKQKQKEYISDDHRSISITALSVMQETVPTPLARHLIEEQNVISITE 448
Db 181 YKKLFAMEFVKYKQKQKEYISDDHRSISITALSVMQETVPTPLARHLIEEQNVISITE 240
QY 449 TLLEVLPEYLDNRNKNFQGYSDKLGKRYAVICDLKYILISKPTTWTERLRMQFLEGFR 508
Db 241 TLLEVLPEYLDNRNKNFQGYSDKLGKRYAVICDLKYILISKPTTWTERLRMQFLEGFR 300
QY 509 SFKLILCTMQGMEIEIRROVQGHIEVDPDWEAAIAIQMLKNILLMFQEWACACDEELLVLA 568
Db 301 SFKLILCTMQGMEIEIRROVQGHIEVDPDWEAAIAIQMLKNILLMFQEWACACDEELLVLA 360
QY 569 YKECHKAVMRCSTSFSSSKTVVQSCGHSLETKSYRVSEDLVSIHLPLSRTLAGLHVRLS 628
Db 361 YKECHKAVMRCSTSFSSSKTVVQSCGHSLETKSYRVSEDLVSIHLPLSRTLAGLHVRLS 420
QY 629 RLGAVSRLHEFSFDFQVEVLVEYPLRCLVLVAQVVAEMWRRNGLSLSQVFFYQDYKC 688
Db 421 RLGAVSRLHEFSFDFQVEVLVEYPLRCLVLVAQVVAEMWRRNGLSLSQVFFYQDYKC 480
QY 689 REEMYDKDIIMLQIGASLMDPNKFLLLVLQRYELAEAFNKTISTKQDQLIKQYNTLIEEM 748
Db 481 REEMYDKDIIMLQIGASLMDPNKFLLLVLQRYELAEAFNKTISTKQDQLIKQYNTLIEEM 540
QY 749 LQVLIYIVGERYVPGVGNVTKKEVTWREIHLICIEPMPHSAIAKNLPENNETGLENY 808
Db 541 LQVLIYIVGERYVPGVGNVTKKEVTWREIHLICIEPMPHSAIAKNLPENNETGLENY 600
QY 809 INKVATFKPGVSGHGYELKDESLEKDFNMFYFHYSKTQHSKAEHQKRRKQENKDEAL 868
Db 601 INKVATFKPGVSGHGYELKDESLEKDFNMFYFHYSKTQHSKAEHQKRRKQENKDEAL 660
QY 869 PPPPPFPCPAFSKVITNLLNCDIMMYILRTVFERRADTDSNLTWTEGMLQWAFHILALGLL 928
Db 661 PPPPPFPCPAFSKVITNLLNCDIMMYILRTVFERRADTDSNLTWTEGMLQWAFHILALGLL 720
QY 929 EEKQQLQKAPEEVTFDYHKASRLGSSAMNIQMLEKLGIPQLEGQKDMITWILQMF 988
Db 721 EEKQQLQKAPEEVTFDYHKASRLGSSAMNIQMLEKLGIPQLEGQKDMITWILQMF 780
QY 989 TVKRLREKSLIVATTSGSESINDEITHDK 1019
Db 781 TVKRLREKSLIVATTSGSESINDEITHDK 811

RESULT 5
AA078576
ID AA078576 standard; Protein; 1400 AA.
XX
AC AA078576;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human protein SEQ ID NO 1238.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN W0200157190-A2.
XX
XX 09-AUG-2001.
XX
```

```
PF 05-FEB-2001; 2001WO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 13-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR WPI; 2001-476283/51.
DR N-PSDB; AAK51709.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
XX Claim 20; Page 3496-3499; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AA080302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 1400 AA;
XX
XX Query Match 36.7%; Score 3359; DB 22; Length 1400;
XX Best Local Similarity 46.4%; Pred. No. 5.8e-276;
XX Matches 659; Conservative 271; Mismatches 425; Indels 64; Gaps 22;
QY 361 MLWDALYKGAARKILHELIFSSFFMEMEYKILFAMEFVKYKQKQKEYISDDHRSISIT 420
Db 1 MLSDKLWKGARSYTHQLFMSLLMDLKYKLLFAVRFAKNTYERLQSDVYDDHDSFSA 60
QY 421 ALSVQMETVPTPLARHLIEEQNVISITETLLEVLPEYLDNRNKNFQGYSDKLGKRYAV 477
Db 61 DLSVQIFVTPSLARMLITEENLMSIITKTFMDHL-RHRDAQGRFQFERYTALQAFKFRV 119
QY 478 YAVICDLKYILISKPTTWTERLRMQFLEGFRSFLKILCTMQGMEIEIRROVQGHIEVDP 537
Db 120 QSLILDLYLVLSKPTWSEDLRQKLEGFDAFLKLLCMQGMDDPITRQVQGHIEPEW 179
QY 538 EAAATAIQMLKNILLMFQEWACACDEELLVLAQVVAEMWRRNGLSLSQVFFYQDYKC 596
Db 180 EAAFTLQMKLTHVISMMDQWCAKDKVLIAYKCLAVLMQCHGYTQGEQITLISICGH 239
QY 597 SLETKSYRVSEDLVSIHLPLSRTLAGLHVRLSRLGAVSRHLHFSFDFQVEVLVEYPLR 656
Db 240 SVETIRYCVSQEKYSIHLPLSVSRLLAGLHVLLSKSEVAYKFPPELLPLSELSPMLIEHPLR 299
QY 657 CLVLVAQVVAEMWRRNGLSLSQVFFYQDYKQREEMYDKDIIMLQIGASLMDPNKFLLLV 716
Db 300 CLVLCAQVHAGMWRNGFSLVNQIYYHNKCRREMFDKDVVYMLQGTGVSMMDPNHFLLM 359
QY 717 LQRELAFAFN-----KTISTK--DODLIKQYNTLIEEMLQVLIYIVGERYVPGVGNVTK 769
Db 360 LSRFELYQIFTPDYGKRFSSEITHKDVVQVQNTLIEEMLYLIIMLGRFSPGVQVNA 419
```


QY	770	E E V T W R I I I H L L C I E P M P H S A I A K N L P E N E N N E T G L E N V I N K V A T F K K P Q V S G H V Y E L K	829
Db	420	T D E I K R E I I H Q L S I K P M A H S E L V K S L P E D E N K E T G M E S V E A V A H F K P G L T G R G M Y E L K	479
QY	830	D E S I K D N M Y F Y H V K T O H S K A B H M K R K Q E N K D E A L P P P P P P P C P A F S K V I N L L N C	869
Db	480	P E C A K E N L F Y H F S R A Q S K A E A Q R K L K R N E D T A L P P V L P P C P L F A S I V N L I Q S	539
QY	890	D I M M Y I L R T V F R A I D T S N L W T E G M L O M A F H A L L G L E E K O O L K A P E E E V T D F D Y H	948
Db	540	D V M L C I M G T I L W A V E H G Y A W S E S M L Q R V L H I G M A L Q E K H L E N V T E H V V T F T Q	599
QY	949	K A S R L G S S A M N --- I O W L E K L G I P Q E C Q K D M I T W I L O M F T V T R L R E K S --- C L I V A T T	1004
Db	600	K I S K P E A P K N S P I I L A M E T L Q N A P Y L E V H K D M I R W I L T F N A V K M R E S S P T S P V A E T	659
QY	1005	S G S E I K N D I T H D K A E R K R A E A A R L H R K I M A Q M S L O K N F I T E T H K L M Y D N T S E M P	1064
Db	660	E G T I M --- E S S R D K K A E R K R A E A R L R R E K I M A Q M S E M Q R H F I D E N K E L F Q O Q L E L D	716
QY	1065	G K E D S I M E E B T P A V S D Y S R I A L G P R K G P S V T E K E V L T C I L C O E E Q E V K T E N N A M V L S A C	1124
Db	717	A S T S A V L D H --- S P V A S D M T I L T A L G P A Q T O V P E Q R Q F V T C I L C O E E Q E V K V E S R A M V L A A F	774
QY	1125	V K S T A L T O H R G K P I E L S G E A L D P L E M D P D L A Y G T Y T G S G H V M A V C W K Q Y E A V O L S S	1184
Db	775	V O R S T V L S K N R S F I Q --- D P E K Y D P L M P H P D L S C G T H T S C G H I I H A C W Q R Y F D S V Q A K E	833
QY	1185	Q Q --- R H V D L D F L E S G E Y L C P L C K S L C N T V I P I P L O P Q K I N S E N A D A L A Q L L T L A R	1239
Db	834	Q R Q Q R L R L T S --- Y D V E N G E I F C L C E C U S I N T V I P L L --- L P P R N I F N R L N --- F S D Q P N L T Q	890
QY	1240	W I O T V L A R I S G Y N I R A K E N P I P I E F N O G M G D S T L E F H S I L S F G V E S S I K Y S N I S I K E M V	1299
Db	891	W I R T I S Q I K A L O F L R K E B S T P N N A S T K N S E N V D E L Q L P E G F R D P R P K I P Y S E S I K E M L	950
QY	1300	I L P A T T Y I R I G L K V P D E R D P R V M L T W S T C A F T I O A I E N L G L D E G K P L F G A L Q N R O H N G	1359
Db	951	T T F G T A T Y K V L K V H N E E D P R V P I M C W S C A V T I Q S I E R I L S D E D K P L F G P L P C R L D D C	1010
QY	1360	L K A L M O F A V A O R I T C P O V L I O K H L V L L S V V L P N I K S E D T P C L L S I D L F H V L G V A I A P P	1419
Db	1011	L R S L T R F A A A H W T V A S V V V Q G H F C P F A S L V P N D S H E E L P C I L D I M F H L L G L V L A F P	1070
QY	1420	S L Y W D D P V D L Q P S S V S S Y N H L Y L F H I T M A H M L Q I L L --- T V D T G L P L A Q V O E D S E A H	1476
Db	1071	A L O C Q D --- F S G I S L G T G D L H F H V T W A H I I Q I L L T S C T E N G M D --- Q E N P P C E E S	1123
QY	1477	S A S S F P A E I S O Y T S G I G C D I P --- G W Y L W S L K N G I T P Y L R C A A L F F H Y L I L G V T P P P E L H T	1535
Db	1124	A V L A L Y K T L H O Y T --- G S A L K E I P S G W L H R S V R A G I M P F L K C S A L F F H Y L N G V S P P D I O V	1182
QY	1536	N S A E G Y S A L C Y L S L P T N L F I L F O E Y W D T V R P L Q R R O A D P A L L N C L K O K N T V R Y P R K	1595
Db	1183	- P G T S H F E L C Y S L S P N L I C L F O N S E T M N S L I E S W C R N S V K Y L E G E R D A I R Y P R E	1241
QY	1596	R N S L I E L P D D Y S C L L N O A S H F C P R S A D D E R H P V L C L F C G A I L C S O N I C C C O I V N G E E V	1655
Db	1242	S N K I L N P E D Y S S L I N O A S N F S C P K S G G D K S R A P T L C L V C G S L L C S Q S Y C C Q T L E G E D V	1301
QY	1656	C A I F H A L H C --- K A R G C A Y P A P Y L D E Y E T D P G L K R G N P L H L	1695
Db	1302	G A C T A H Y S C G S G V G I F L R V R E C O V L F L A G K T K C F Y S P Y L D D Y G E T O G L R R G N P L H L	1361
QY	1696	S R E R Y K R L H V W O O H C I I E I A R S Q E T N O M L F G N W O L L	1734
Db	1362	C K E R F K I Q K L W H O S V T E E I G H A Q E A N Q T L V G I D W O H L	1400

RESULT 6

AAM79560

ID AAM79560 standard; Protein: 1400 AA.

RESULT 6
AAM79560
ID AAM

XX	AAW79560;	
XX	06-NOV-2001 (first entry)	
XX	Human protein SEQ ID NO 3206.	
XX	Human; cytokine; cell proliferation; cell differentiation; gene therapy;	
XX	vaccine; peptide therapy; stem cell growth factor; haematopoiesis;	
XX	tissue growth factor; immunomodulatory; cancer; leukaemia;	
XX	nervous system disorder; arthritis; inflammation.	
OS	Homo sapiens.	
XX	WO200157190-A2.	
XX	09-AUG-2001.	
XX	05-FEB-2001; 2001WO-US04098.	
XX	03-FEB-2000; 2000US-0496914.	
XX	27-APR-2000; 2000US-0560875.	
XX	20-JUN-2000; 2000US-0598075.	
XX	19-JUL-2000; 2000US-0620325.	
XX	01-SEP-2000; 2000US-0654936.	
XX	15-SEP-2000; 2000US-0663561.	
XX	20-OCT-2000; 2000US-0693325.	
XX	30-NOV-2000; 2000US-0728422.	
XX	(HYSE-) HYSEQ INC.	
XX	Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;	
XX	Zhao Q, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;	
XX	Xue AJ, Yang Y, Wejhrman T, Goodrich R;	
XX	WPI; 2001-476283/51.	
XX	N-PSDB; AAK52693.	
XX	Nucleic acids encoding polypeptides with cytokine-like activities,	
XX	useful in diagnosis and gene therapy	
XX	Claim 20; Page 284-285; 6221pp; English.	
XX	The invention relates to polynucleotides (AAK51456-AAK53435) and the	
XX	encoded polypeptides (AAM78323-AAW80302) that exhibit activity elating to	
XX	cytokine, cell proliferation or cell differentiation or which may induce	
XX	production of other cytokines in other cell populations. The	
XX	polynucleotides and polypeptides are useful in gene therapy, vaccines or	
XX	peptide therapy. The polypeptides have various cytokine-like activities,	
XX	e.g. stem cell growth factor activity, haematopoiesis regulating	
XX	activity, tissue growth factor activity, immunomodulatory activity and	
XX	activin/inhibin activity and may be useful in the diagnosis and/or	
XX	treatment of cancer, leukaemia, nervous system disorders, arthritis and	
XX	inflammation.	
XX	Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666	
XX	(AAM80020) are omitted as the relevant pages from the sequence listing	
XX	were missing at the time of publication.	
XX	Sequence 1400 AA;	
XX	Query Match 36.7%; Score 3356; DB 22; Length 1400;	
XX	Best Local Similarity 46.2%; Pred. No. 1e-275;	
XX	Matches 659; Conservative 272; Mismatches 424; Indels 70; Gaps 2	
Qy	355 CLISRLMLWDALYKGAKKILHELIFSSFFMEMEYKKLFAMEFVKYKQLQKEYISDDHD 410	
Db	1 CLV-----DSKLMKGARSVYHQLFMSLSLLMDLYKKKLFARFAKNYERLQSDYVTTDDHD 54	
Qy	415 RSISITALSVOMFTVPTIARHLHIEQNVISVITETLLEVLPEYLDNRNKNFQGYG---Q 471	
Db	55 REFVSADLSVQIFTVPISARMLITENLMSIIKTFMDHL-RHRDQAQGRFQFERYALQA 113	
Qv	472 DKLGRVYAVICDLKYLISKPTIWTFLRMOLFGRFSFLKILTCMOGMEEIRVOGSHI 531	


```
XX PF 11-JAN-1999; 99US-0228317.
XX PF
XX PR 02-DEC-1997; 97US-0982956.
XX PA (CALY ) CALIFORNIA INST OF TECHNOLOGY.
XX PI Kwon YT, Varshavsky A;
XX WPI; 2001-090278/10.
XX DR N-PSDB; AAC86934.
XX PT Inhibiting the N-end rule pathway in mammalian cells for treating
XX PT infections and various diseases associated with muscle tissue wasting,
XX PT by inhibiting the expression of Ub1 gene
XX XX
XX PS Example; Column 27-30; 18pp; English.
XX CC
XX CC The present sequence represents a partial Ub1 enzyme. Ub1 is an E3-type
XX CC protein of the ubiquitin system. Specifically, it is a ubiquitin-protein
XX CC ligase. The enzyme is specific for destabilising residues exposed at
XX CC the N-terminus of protein substrates. Inhibition of the expression of
XX CC Ub1 gene in a cell results in inhibition of the N-end rule pathway.
XX CC The method is used for treatment of mammalian cells infected with an
XX CC intracellular pathogen, e.g. Lysteria monocytogenes or Yersinia
XX CC enterocolitica. Inhibition of N-end rule pathway is also useful for
XX CC treating various diseases associated with wasting of muscle tissue and
XX CC infections.
XX CC
XX CC Sequence 333 AA;
XX CC
XX CC Query Match 18.8%; Score 1716; DB 22; Length 333;
XX CC Best Local Similarity 99.7%; Pred. No. 4.8e-137;
XX CC Matches 332; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX CC
QY 707 MDPNKFLLVLRQYELAEAFNKTISTKDDLIKQYNTLIEMLQVLIYIGERYVPGVN 766
DB 1 MDPNKFLLVLRQYELAEAFNKTISTKDDLIKQYNTLIEMLQVLIYIGERYVPGVN 60
QY 767 VTREVTMREIHLCTEPMHSAIAKNLPENNETGLENVINKVATFKPGVSGHGV 826
DB 61 VTREVTMREIHLCTEPMHSAIAKNLPENNETGLENVINKVATFKPGVSGHGV 120
QY 827 ELKDESLEKDFNMYHYHYSKTOHSAEHMQKRRKQENKDEALPPPPPPFCFAFSKVINL 886
DB 121 ELKDESLEKDFNMYHYHYSKTOHSAEHMQKRRKQENKDEALPPPPPPFCFAFSKVINL 180
QY 887 LNCDDIMMYILTVTFERAIDTDSNLWTEGMLQMAFHIALGLEEKQQLQKAPEEVTFDF 946
DB 181 LNCDDIMMYILTVTFERAIDTDSNLWTEGMLQMAFHIALGLEEKQQLQKAPEEVTFDF 240
QY 947 YHKASRLGSSAMNTOMLLEKLGIPQLEGQKDMITWILQMDFTVKRLREKSLIVATTSG 1006
DB 241 YHKASRLGSSAMNTOMLLEKLGIPQLEGQKDMITWILQMDFTVKRLREKSLIVATTSG 300
QY 1007 SESIKNDEITHDKEAERKKAERAAARLHRQKIM 1039
DB 301 SESIKNDEITHDKEAERKKAERAAARLHRQKIM 333
RESULT 12
ABB90168
ID ABB90168 standard; Protein; 452 AA.
XX AC ABB90168;
XX XX
XX DT 24-MAY-2002 (first entry)
XX XX
XX DE Human polypeptide SEQ ID NO 2544.
XX CC
XX KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
XX KW antifungal; hepatotropic; antidiabetic; antiinflammatory; antitumor;
XX KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
```

```
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein.
XX XX
OS Homo sapiens.
XX PN WO200190304-A2.
XX PD 29-NOV-2001.
XX 18-MAY-2001; 2001WO-US16450.
XX 19-MAY-2000; 2000US-205515P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Birse CE, Rosen CA;
XX WPI; 2002-122018/16.
XX N-PSDB; ABL90577.
XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
XX prevention of neural, immune system, muscular, reproductive,
XX gastrointestinal, pulmonary, cardiovascular, renal and proliferative
XX disorders -
XX Claim 11; SEQ ID NO 2544; 208lpp + Sequence Listing; English.
XX The invention relates to novel genes (ABL89449-ABL90853) and proteins
XX (ABB9040-ABB9044) useful for preventing, treating or ameliorating
XX medical conditions e.g. by protein or gene therapy. The genes are
XX isolated from a range of human tissues disclosed in the specification.
XX The nucleic acids, proteins, antibodies and (ant)agonists are useful
XX in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
XX and ovarian cancer and other cancers of the adrenal gland, bone, bone
XX marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
XX (b) immune disorders e.g. Addison's disease, allergies, autoimmune
XX haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
XX disease, multiple sclerosis, rheumatoid arthritis and ulcerative
XX colitis; (c) cardiovascular disorders such as myocardial ischaemias;
XX (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
XX epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
XX and parasitic infections.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX CC
XX CC Sequence 452 AA;
XX CC
XX CC Query Match 11.7%; Score 1071.5; DB 23; Length 452;
XX CC Best Local Similarity 44.9%; Pred. No. 6.4e-82;
XX CC Matches 207; Conservative 79; Mismatches 142; Indels 33; Gaps 7;
XX CC
QY 1298 MVILFATTYIRGLKVPDPDRPRVPMLTWSTCAFTQAIENLLGDEKPLFGALQNRQH 1357
DB 1 MLETFGTATYKVGKLVHPNEEDPRVPMCMGSCAYTIQSIERLSDEKPLFGPLCRLD 60
QY 1358 NGIKALMOFAVAGRITCPQVLQKHLVRLLSVVLPNKSBDTPCLLSIDLFLVLGAVLA 1417
DB 61 DCLSLTRFAAAHWTVASVSVQVQGEKFLASLPNDSDHEELPCILIDIMFHLVLGLVLA 120
QY 1418 FPSLYWDDPVDLQPSVSSSYNHLFLHITMAHMLQILL---TVDTGLPLAQVQDESEE 1474
DB 121 FPALQCD-----FSGISLGTGDLHIFHLVTHMAHIIQLITSCTEENGMD--QENPPCEE 173
QY 1475 AHSASSFFAEISQYTSGSGCCDIP-CWYLMVSLKNGITPYLRCAALFFHLLGYTPPEEL 1533
DB 174 ESAVLALYKTLHQYT-GSALKETPSGWHLWRSVRAGIMPFKCSALFFHYLNGVPSPPDI 232
QY 1534 HTNSABGEVSALCSYLSLPTNLFLLFQYWDTVRPLLRORCADPALLNCLKQNTVVRYP 1593
DB 233 QV-PGTSHEHLCSYLSLNNLCLFQENSEINLSIESCRNSEVSKRYLEGERDAIRYP 291
QY 1594 RKRNLSIELPDDYSCLLNQASHFRCPRSADDERKHPVLCFLFCGAILCSQNICQEIYNGE 1653
```



```

Db      292 RESNKLINLPEDYSSLIQASNFSPKSGGDSRAPTLCLVCGSLCSQSYCCQTELEGE 351
QY      1654 EVGACIFHALHC-----KARGCAYPAPYLDYGETDPGLKRGNPL 1693
Db      352 DVGACTAHTYSCGSGVGIFLRVREQVLFLAGTKGCFSPYLDYGETDQGLRRGNPL 411
QY      1694 HLSRERYKHLVWQOHCIIIEIARSQETNQMFLFGFNWQLL 1734
Db      412 HLCRERFKKIOKLWHOHSVTBEIGHAQEANTLVGIDWQHL 452

```

RESULT 13

ABG05917 standard; Protein; 487 AA.

XX ABG05917;

XX 13-FEB-2002 (first entry)

XX Novel human diagnostic protein #5908.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX N-PSDB; AAS70104.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

XX Claim 20; SEQ ID No 36276; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 487 AA;

Query Match 9.6%; Score 877; DB 22; Length 487;
Best Local Similarity 59.2%; Pred. No. 2.7e-65;
Matches 189; Conservative 0; Mismatches 0; Indels 130; Gaps 2;

QY 643 EDFQVEVLVEYPLRCLVLVAQVVAEMRRNGLSLSISQVFFYQDYKCREMYDKDIIML-- 700

Db 3 EDFQVEVLVEYPLRCLVLVAQVVAEMRRNGLSLSISQVFFYQDYKCREMYDKDIIMLQK 62

QY 701 -----QIGASLMDPNKFLLLVLQRYELAAEFNKITSTKDQDLIKQYNTLIEMLQVL 752

Db 63 KYDAGFLSQIGASLMDPNKFLLLVLQRYELAAEFNKITSTKDQDLIKQYNTLIEMLQVL 122

QY 753 IYIVGERYPVGNGVTKEEVTMRIIHLCTECPHPSAIAKNLPENNETGLENVINKV 812

Db 123 IYIVGERYPVGNGVTKEEVTMRIIHLCTECPHPSAIAKNLPE----- 168

QY 813 ATEKKPGVSGHGVYELKDESILKDFNMYFYHYSKQHSKAEHMQKRRKQENKDEALPPPP 872

Db 169 ----- 168

QY 873 PPEFCPAFSKVINLLNCDINMYILRTVFERAIDTDSNLWTEGMLQMAFHILALGLEEKQ 932

Db 169 -----AFHILALGLEEKQ 182

QY 933 QLQKAPEEEVTFDFYHKAS 951

Db 183 QLQKAPEEEVTFDFYHKAS 201

RESULT 14

ABG05881

ID ABG05881 standard; Protein; 258 AA.

XX ABG05881;

XX 13-FEB-2002 (first entry)

XX Novel human diagnostic protein #5872.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX N-PSDB; AAS70068.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

XX Claim 20; SEQ ID No 36240; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

xx SQ Sequence 258 AA;
 Query Match 8.6%; Score 790; DB 22; Length 258;
 Best Local Similarity 100.0%; Pred. No. 2.4e-58;
 Matches 149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1076 TPAVDYSRIALGPKRGPSVTEKEVLTCILCOEBOEVKIIENNAVLISACVQKSTALTOHR 1135
 DB 109 TPAVDYSRIALGPKRGPSVTEKEVLTCILCOEBOEVKIIENNAVLISACVQKSTALTOHR 168
 QY 1136 GKPIELSGEALDPLFMDPLAYGTYTGCGHVMHVCWQKYFEAVQLSSQORIHVDLFDL 1195
 DB 169 GKPIELSGEALDPLFMDPLAYGTYTGCGHVMHVCWQKYFEAVQLSSQORIHVDLFDL 228
 QY 1196 ESGEYLCPLCKSLCNTVIPIIPIQPKIN 1224
 DB 229 ESGEYLCPLCKSLCNTVIPIIPIQPKIN 457

RESULT 15
 ABB93333
 ID ABB93333 standard; Protein; 1225 AA.
 XX AC ABB93333;
 AC ABB93333;

DT 31-MAY-2002 (first entry)
 XX Herbicidally active polypeptide SEQ ID NO 2544.
 DE Herbicidal; plant; agriculture; herbicide.
 KW Arabidopsis thaliana.
 XX OS Arabidopsis thaliana.
 XX PN W0200210210-A2.
 XX PD 07-FEB-2002.
 XX PF 28-AUG-2001; 2001WO-EP09892.
 XX PR 28-AUG-2001; 2001WO-EP09892.
 XX PA (FARB) BAYER AG.
 XX PI Tietjen K, Weidler M;
 XX PT WPI; 2002-269010/31.
 XX DR Identifying plant target proteins for herbicidally active compounds,
 PT comprising aligning and comparing nucleic acid or amino acid sequences
 PT from plant with nucleic acid or amino acid sequences from non-plant
 PT organisms -
 XX PS Claim 5; SEQ ID NO 2544; 261pp + Sequence Listing; English.
 XX CC The invention relates to identifying target proteins
 CC (ABB90790-ABB94016) for herbicidally active compounds, comprising

CC aligning and comparing nucleic acid or amino acid sequences from plant
 CC with nucleic acid or amino acid sequences from non-plant organisms using
 CC suitable search parameters, where plant sequences having an E-value
 CC greater by a factor of 3 than the E-value of most similar non-plant
 CC sequences are selected. The polypeptides or nucleic acids encoding them
 CC are useful for identifying modulators. The identified modulators are
 CC useful as herbicides.

xx SQ Sequence 1225 AA;

Query Match 6.6%; Score 606.5; DB 23; Length 1225;
 Best Local Similarity 21.1%; Pred. No. 1.4e-41;
 Matches 275; Conservative 190; Mismatches 422; Indels 417; Gaps 47;

QY 60 MDPDLKO-EESVQMSIFTPLEWYLFGEDEPDICLKHSGAFQ--LCGRVFKSGTYS 116
 DB 79 VSPMKKRRESNMN-----LQWLFDQDEPVSRLNKLAKLNDLQRCVCGSVQNGDIAVR 133
 QY 117 CRDCAIDPTCVLCMDCFQDSVHKHRYKMHSTGGGFCDCGTAEAKTGPFCVNHPEGRA 176
 DB 134 CRTCEPDPTCAICVPCFONGDHNHDSYI-IYTGCGCCDCGTAEAKTGPFCVNHPEGRA 189
 QY 177 GTIKENSRCPLEEVIVQARKIFPSVIVVEMTIWEEKELPPELQIREKNERYVCVLF 236
 DB 190 ---SEQIR-PISENLANSVGPILDAL-----FTCNKK----- 218
 QY 237 NDEHSHYDRVYSLQALDCELAELHTAIDKEGRRAVKAG-AYAQCEAKEDIKSHS 295
 DB 219 -----LLSAESSGQKCARSDNTLVLQKSNELTF-- 248
 QY 296 ENVSOHPHVEVLHSEIMAHOKFAURLGSMWNKIMSYSDSFROIQCOACLRPEPSENPC 355
 DB 249 -----IVVEMLEFMSSES--LSFVSRRISSG----- 277
 QY 356 LISRLMLDAKLYKGARKILHELIFSSPFMEYKLFAMEFVYKYLQKEVYISDDHHR 415
 DB 278 LLSILLKAEERLDQVMKMLHDL-FLKLIGDPVFKCEFAKAVSYVPVIVSEVVGQGTGN 336
 QY 416 SIS-----ITALSVQMTVPTLARHLIEONVISVITETLLLEVLPEYLDNRNKNFQY 471
 DB 337 AFKKYPLLSTFSVQILVPTLTPFLVKEMNLLAMLLGCLSLDFVSCSGEDGLL-----QA 391
 QY 472 DKLGRV---YAVICDLXYI---LISKPTIWRERMOFLGFFSFLKILTCMQGMEI 523
 DB 392 TKLERLCETSERVIGDLKFMVSHAIVSKYATHEHR-----ELSRWLTLLTFAQGMNPL 445
 QY 524 RROVQGHTEVDPDW-----EAAIAIQMLKNILLMFQEWACAD- 561
 DB 446 KRETGIPIDEENDYMHLEFVLGHSTAVIHSLLVNGTYSAADEEINDRNAKEEFKCDG 505
 QY 562 -----EELLVA-----YKECHKAV 576
 DB 506 DGERYAKVGRLSHEDSVCTAIVSSSFDSSMASEVHKIDPFHALLPSSAIYLIRECKVL 565
 QY 577 MRC-----STFSISSKT----- 589
 DB 566 ETCLENGEGISKFLCKLSSSGRNIPESKMSWPRDLNVTGSGVSSNLASSRDPSTG 625
 QY 590 -----VVO-----SCGHSLETKS----- 602
 DB 626 LSPLCGDTQTNLSLDNVCVGPVGVQTDVTADSKRVSCNSADLTKNASGLRILGLCDWPD 685
 QY 603 -YRVSEDLVSIHLPLSRTLGLHVRSLR-----GAVSRLHEFVSPFEDFOVEL--- 650
 DB 686 HYDVSSQALSVHPLHRLLSLLIQALRICYGESASVNGYSISHE-IPRADFPSSVIGDF 744
 QY 651 -----VEYPLRCLVLVAQVAMRMNRNGLSLISQVIFYQDVKCR--EEMYDKDIIM 699
 DB 745 HPCGFSALVMEHVLQIRVCAQVIAGMKKNGDSAL-----VSCSEQGLELDLFL 796
 QY 700 LQIGASLMDPNKFLLLVQRYELAEAFNKTISTKQDLIKQY--NYLIEMLQVLIYIVG 757
 DB 797 LQCCAALAPADSYVDKLLSRFGLS-----SYLSLNPDTINEXYVPIVLLMLLGLLIQILQ 851

Qy	758	ERYVPGVNTKEEVMTREIHLICIEPMHPSAIAKNLPENNETGLENVINKVATFKK	817
Db	852	ERRFCGLSTA---ESLRREIIFKLATGFTHSQLVKSLPRDLKSDLEQVELDVSVYCN	908
Qy	818	PVSGHGVYELKDESLDNMYFYHYSKTQHSKAEHMQKRRKQENKDEALPPPPPEFC	877
Db	909	PSGMNQGKYSLOSSCWKELDLY-----HPRWQSRDLQSAERF-----SRYC	950
Qy	878	-----PAFSVINLLNCDIMMYILRTVFERAIDTDSNLWT---EGMLQOM	918
Db	951	GVSAITTLQLPWRMIYPLKGLARIGTKATFQIISSALYVALQSGTSVKSRAPDGLIT	1010
Qy	919	AFHIALGLLEEKQOLQKAPEE---EVTDFYHKAS-----RLGSSAMNIOMLLEKLG	969
Db	1011	ALQLLSLSLDICTOORQNSQDCCLENSIPILELAGLEIIGIAOQTERESLILSLVSLMK	1070
Qy	970	IPQLEGOK-----DMITWI---LQMEDIV---KRLREKSCLIIVATTSGSESKNDE	1014
Db	1071	TRMGDRHQFPPEPGSCNIISSWIGNLLKFSAIDSVCMNLLQSLAPEVVGQSGFDKVMGS	1130
Qy	1015	ITHDKAEAKRKAERLHRQKIMAOMSALQKNFIETHKLMYDNTSEMPGKE---DSI	1070
Db	1131	TSDEKRRKAKERQAA-----IMAKMKAQSKFSLTLSSMDD--DDPRSEFETS	1181
Qy	1071	MEESTPAVSDYSRIALGP-KRGP-----SVTEKEVLTC	1103
Db	1182	MEHDSIAVREVCSILCHDPDSKDPVSFLIFLQVGSMTDAIIOCDC	1225

Search completed: September 25, 2003, 14:54:33
Job time : 111 secs

Result No.	Score	Query Match	Length	DB	ID	Description
1	176	1.9	2662	4	US-09-595-684B-31	Sequence 31, Appli
2	172	1.9	3878	4	US-09-914-259-11	Sequence 11, Appli
3	156.5	1.7	2710	1	US-08-480-604A-6	Sequence 6, Appli
4	156.5	1.7	2710	2	US-08-405-496A-6	Sequence 6, Appli
5	156.5	1.7	2710	3	US-08-915-136-B	Sequence 6, Appli
6	156.5	1.7	2710	4	US-08-957-310-6	Sequence 6, Appli
7	156.5	1.7	2710	4	US-10-011-366-B	Sequence 6, Appli
8	156	1.7	3433	4	US-09-091-501B-10	Sequence 10, Appli
9	150.5	1.6	1388	4	US-09-572-191-2	Sequence 2, Appli
10	150.5	1.6	1388	4	US-09-723-262-2	Sequence 2, Appli
11	150.5	1.6	1388	4	US-09-723-219-2	Sequence 2, Appli
12	147.5	1.6	3248	1	US-08-353-700-1	Sequence 1, Appli
13	147.5	1.6	3248	5	PCI-US95-16216-1	Sequence 1, Appli
14	140.5	1.5	959	4	US-09-914-259-67	Sequence 67, Appli
15	140.5	1.5	1708	1	US-08-493-092-2	Sequence 2, Appli
16	140.5	1.5	1708	1	US-08-508-836A-2	Sequence 2, Appli
17	139.5	1.5	2482	1	US-08-328-254-6	Sequence 6, Appli
18	139	1.5	976	3	US-09-104-324B-4	Sequence 4, Appli
19	137	1.5	2210	4	US-09-309-572-7	Sequence 7, Appli
20	134	1.5	1242	4	US-09-107-532A-5241	Sequence 5241, Appli
21	133.5	1.5	961	1	US-09-914-259-66	Sequence 66, Appli
22	133.5	1.5	3056	1	US-08-508-836A-8	Sequence 8, Appli
23	133.5	1.5	3056	2	US-08-629-001A-3	Sequence 3, Appli
24	133.5	1.5	3056	2	US-08-874-266-2	Sequence 2, Appli
25	133.5	1.5	3056	3	US-08-642-247D-3	Sequence 3, Appli
26	133.5	1.5	3056	3	US-08-952-127-3	Sequence 3, Appli
27	133.5	1.5	3056	3	US-08-952-014C-3	Sequence 3, Appli

```
Qy 437 --IEQNVISVITETLLEVLPE--YLDNNKFNFGYSODKLGRIYAVICDLKYLISK- 491
Db 971 ESLKHQETINTLKSISEEYRNLMHEENTGETKDEFQKMWG-----IDKQDLEAKN 1024
Qy 492 -PIWTRLRMQFLEGRPSFLKILTCMQGMEETRRQVGHIEVDPDWEAAIAIOMQLK-- 548
Db 1025 TOTLTADVKNDEIEEQOR--KIFSLTOEKNEQLQML-----ESVIAEKELQKTD 1071
Qy 549 ---NILLMFOBCACDELLIVA--KECHKAYMRCSTSFSSSKTVVQSCG-----HS 597
Db 1072 LKENIEMTIEN---QBELRLGDELKQOEIVAOEKNHAIKKEGELSRTCDRLAEEVK 1127
Qy 598 LETKSVRSE---DLVSIHPLS-----RTLGLHRLSLRGAVSRL 636
Db 1128 LKESQOOLEKQOOLLNVQEMSEMOKKINEIENKLNELNKTLEHMETERLELAQKL 1187
Qy 637 HEFVSFEDFOVELVVEPLRCLVLVAQVVAEMRRNGLSISVYFYQDVKCREMYDKD 696
Db 1188 NE--NYEE-----VKSITKE--RKVLKELQKSFETERDHLRGVIREIE 1236
Qy 697 IIMLOIGASLMDPNKFLILLVQRY-ELAAEAFNKTISTK-----DQDLIKQYNTLIEML 749
Db 1227 ATGLQTKREL---KIAIHLKHEQETIDELRRSVSEKTAQIINTQDLEKSHTKLOEE-- 1280
Qy 750 QVLIYIVGRIYVPGVGVNVTKEETMRIIHL-----LCIEPMPSAIAKNLPEN 798
Db 1281 -IPVLHEEQELLNVKVKSTQETMNELELLTEOSTTKDSTTLARIEMERLRLNEKFOES 1339
Qy 799 ENNETGLNIVNKVATPKPGVSGHYVELKDSKDFNMVYFH---YKSTQHSKAEHM 854
Db 1340 QEEIKSLTKERDNLTKIE-----ALEVKHDLKE-----HIRETLAKIQESQSKQE 1386
Qy 855 QKRRKQENKDEALPPPPPEFOPAFSKVINLNCDIMMIL-----RTV----- 899
Db 1387 QSLNMKEKNETTKIVSEMEQFKPKOSA---LLRIEIEMLGLSKRLQESHDEMSVAKE 1443
Qy 900 -----FRADTDSNLWTEGMLQ-WAPHILALGLLEKEQLOKA-----PEEVEYTFDYH 948
Db 1444 DDLQRLQEVLSQSDQKQENIKETIAVAKH-----LETEELKVAHCLCKEOEETINELR 1497
Qy 949 KASRLGSSAMNIOMLE-----KLKGIPQLEGQDMITWILQMDTVKRLREKSLI 1000
Db 1498 NLSEKETEISTIQOLQEAINDKLNKQIQEYIEKEEQLN-IQISEVQENVNELQKFEHR 1556
Qy 1001 VATTSGESIKND--EITHD-----KEAPKRKAEEAARLHR-----OKIM 1039
Db 1557 KAKDSALQSTESKMLELTNRLQESQEBIQIMKEKEEMKRVQOEALQIERDQKENTKEIV 1616
Qy 1040 AQMSALQ-----KNFIETHKL-----MYDN 1059
Db 1617 AKMKEQKEYQFLKMTAVANETOEKMCEIEHLKQEFQETQKLNLENIETENIRLQIILHEN 1676
Qy 1060 TSEMPG---KEDSIMEESTPAVSDYSRIALGPKRGPSTVEKVELVCILOEQEYKVI- 1114
Db 1677 LEEMSVTKRDDRDSRVEETLKVERDQKCNL-----RETITRDL-EKQEBELKIV 1725
Qy 1115 -----ENNAMY--LSACVQKSTALTOHRGKPIELSGEAL 1146
Db 1726 HMHLKHEQETIDKLRGIVSEKTEINISNMQDLHNSDAL 1764
```

RESULT 2

```
US-09-914-259-11
; Sequence 11, Application US/09914259
; Patent No. 6495336
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/09/914, 259
; CURRENT FILING DATE: 2000-11-21
```

```
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 3878
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-914-259-11

Query Match      1.9%; Score 172; DB 4; Length 3878;
Best Local Similarity 19.3%; Pred. No. 4.7e-05;
Matches 249; Conservative 188; Mismatches 409; Indels 446; Gaps 61;

Qy 187 LNEEVIVCARFIPSVIKVYVEMTWEEKEKELPPB-LQIREKNERVYCVLFENDEHHSYDH 245
Db 908 LNEELHLQ-RINPTTVK--MKSSVFDEBKTFVATLEMGVEVVEKOTTELMKEKLEVTKRE 963
Qy 246 VIYSLQALDCELABAOLH--TTAIDKEGR--RAYKAGAYAACQAKEDIKSH--SENVYS 299
Db 964 KLELSQRLSDLSQKQKHGEISPLNEEVKSLQKEQVSLRCREL-EIINHNAENVQ 1022
Qy 300 QHPLHVEVL-----HSEIMAHOKFALRLGSMWNKIMSYS 334
Db 1023 SCDTQVSSLLDGVVTMTSRGAEVSGSVKNKSFGEESKIMVEDKV-----SPENMTVGEES 1077
Qy 335 DFRQIFC-----QACLREEPDSENCLISRLMLW-----DAKLYKGARKILHELIES 381
Db 1078 KOEQILDLHPLSVTKESSLRATQPSNDKLOKELNVLSEQNDLRLQMEAQRICLSLVYS 1137
Qy 382 SF-----FMEMEYKK-----LFAMEFVKYKQLOKEYSIDHDHRSISITALSVO- 425
Db 1138 THVDQVREYMEKDKALCSLKEELIFAOE-EKIKELQKIH-----QLELOTKMQOE 1188
Qy 426 -----MFTVPTLARHLIEBQNVISVITETLLEVLPEYL-----DRNNKFN 466
Db 1189 TGDEGKPLHLIGLKQKAVSEE--CSYFLOQLCSVLGEYTPPALCKEVAEDKENSQDY 1245
Qy 467 QYGSODKLGRIYAV-ICDLK-----YILISKPT-----IW---TERLRMQFL 504
Db 1246 ISENEPELDQYRYEVQDFQENMHTLLNKVTEYNKLLVLQTLRLSKQQDGMKLEFG 1305
Qy 505 E-----GFRSFLKIITCMOGMEIRRVQGHIEVDPDWEAAIAIOMKLNILLMFQEMCA 559
Db 1306 EENLPKETEFLSIHQMTNLEDI-----DVNHKSKLSSLODLEKTKLEEGV--- 1352
Qy 560 CDEELLVAYKECHKAVMRCSTSFSSSKTVVQSCG-----HSLKTSYRVSEDLVSI 612
Db 1353 --QEL-----ESLISSLOQQLKETEYAEITHCLQKRLQAVSESTVPP 1394
Qy 613 HLPL-----SRTL-AGLHVRLSRLGAVSRHLHEFVSFEDFQVEVLVE--YPLRCLV 659
Db 1395 SLPVDSVVITESDAQRTMYPGSCVRKKNIDGTTFESGEFGVKEETNIVKLEKQYQSOLEE 1454
Qy 660 LVAQVVAEM---WRRNGLSLIS-----OV 680
Db 1455 EVAKVIVMSIAFAQQOTELSRISGGKENTASSQAHAVCQOQHIFNEMKLSODQIGFOT 1514
Qy 681 FYYQDVKCR-----EEMYDKDIIML--- 700
Db 1515 FETVDVKFEKPKPLSKELGEHKGKELLNSDPHDIPESKDCVLTISEEMFSKDKTFIVR 1574
Qy 701 -----QIGASLMDPNKFLILLVQRYELAEAFNKTISTKQDOLIKQY----- 741
Db 1575 QSIHDEISVSSMDASRLMLNEEQLE-----DMRQELVRYQOEHQQAQTORSSIDN 1624
Qy 742 NTLIEEMLOVLIYVIGERVVPGVGNVTKEETMRIIHL-----LCIEPMPSAIAKNLP 796
Db 1625 ENLVSESEREVLL-----BELEALKOLSLAGREKUCCELNRNSTQTQNGN 1668
Qy 797 ENE-----NNETGLNIVNKV-----ATFFKPKGVS 821
Db 1669 ENQGEVEEQTFKEKELDRKPEDVPPPEILSNERYALQKANNRLKILLEVTKTTAAVEETI 1728
Qy 822 GHGVVELKDESUKDFNMVYHYSKTO-----HSAEHMQKRRKQEN---KDEALP--- 869
```

Db 1729 GRVLGILDRS-----SKSQSSASLIWRSEAEASVSKVHEEHTRVTDPSIPS 1777
QY 870 --PPPPPEPCPAFSKV-----INLNCIDIMYI---LRTVFERAI 904
Db 1778 SGSDMPRNDINWMSKVTEEGTSLQRLVRSFGAGTEIDPENEELMLNISRLQAAYEKLL 1837
QY 905 DFDNLWTEGMLQMAFHILALGLEK-QOOLQKAP-----EVTDFYHKASRLGSSAMN 959
Db 1838 EASE--TSSQLEHA-KVTQTELMRESFROKQEAATESLKQCELRRLHEESR---AREQ 1891
QY 960 IOMLLEKLGIPQEGQDKMIT-----WILQMFDTVKRLREKSLIIVATTSGSP-SIKNDE 1014
Db 1892 LAVELSKAEGV--IDGVADEKTLFEROIQEKTIDIRLQE--LLCASNRLQELAEQQQ 1947
QY 1015 ITHDKEKAERK--KAERARLHROKIMQAOMALQKNFIE----THKLMYDNTSEMPGKE 1067
Db 1948 IQEERELLSROKAMKAEGVP-EQOLLOETELKMKLEVOQCAEKVRDDDLQKQVKALE 2006
QY 1068 DSI-----MEESTPAVSDYSRIALGPKRGPVTEKEVLTCTILCOEEQEVKIENNA 1118
Db 2007 IDVEQVSRFIELEQEKNTFELMDL-----RQOQNALEKQLEKMRKFLDEQAIDREHER 2059
QY 1119 MYLSACVOK-----STALTQHRGKPIE 1140
Db 2060 DVFOEQIEQKLEQQLKVVPREFQISEHOTREVE 2091

RESULT 3
US-08-480-604A-6
; Sequence 6, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,604A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLTA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPND-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-480-604A-6

Query Match 1.7%, Score 156.5; DB 1; Length 2710;
Best Local Similarity 18.6%; Pred. No. 0.00065;
Matches 189; Conservative 154; Mismatches 354; Indels 317; Gaps 44;

QY 288 KEDIKSHSNVSOHPLHVEV---LHSEIMAHQKFAIRLGSWMNMKIMSYSSDFRQIFCOAC 344
Db 355 KSEIFSKLENLVSDLEIKIAFALGSI--NOALISKOGSYLTNLVIEQVKNRYQFLNOH 412
QY 345 LREPDSENPCLISRLMLDAKLYKGARKILHELIFSSFFPMEMEKYKLFAMEFYKYYKQL 404
Db 413 LNPAIESDNN-----PTDTTKTFHDSLFNSATAE---NSMFLTKIAPY---L 453
QY 405 QKEYISDHDHRSISITA-----LSVQMTV-PTL-ARHLIE---EONVISVITE 448
Db 454 QVGFMPARS-TISLSPGAYASAYDFINLOENTIEKTLKASDLIEKFFENNLSQLTE 512
QY 449 TLLEVLPEYLDNRNKNFQYSDQKLGRVYA-----VICDLKYLISK-PTIWT 497
Db 513 QEINSLWSPQASAKYQFEKYVRDYTGSGLSSENGVDNFKNTALDKNVLLNKNIPSNVVE 572
QY 498 RLMOQFLEGRSFLKILTCMOGMBEIRROVGOHIEVDPDWEAAITAQMLKNILLMFOEW 557
Db 573 E-----AGSKNYVHYIQLQGGDISYEATCNLFKNP--KNSIIQRNMNESAKSY--F 622
QY 558 CACDEELLVAYKECHKAVMRCSTSFISSSKTVVQSCGH-----SLETKSYR 604
Db 623 LSDDGESIL-----ELNK--YRIPRLKNKEKVKVTFIIGHKDEPNTSEFARLSVDSLSNE 676
QY 605 VSEDLVSIHLPLSRTLGLVRLSRLGAVSRHLFEVDFEQVEVLVYPLRCLVLAQV 664
Db 677 ISSFLDTIKLDISPK---NVEVNLGC-----NMFYSY-DFNVE--ETYPCKLLLSIMDK 724
QY 665 VA-----EMWRNGLSLISQVFFYQDVVKCREMY--- 693
Db 725 ITSTLPDVNKNSTIGANQVEVRINSEGRKELLASHGKWINKEEAIMSDLSKEYIFPDS 784
QY 694 -----DKDILMQLGASLMDPNKELLVLQ-----RYELAE 724
Db 785 IDNKLKAKSNIPCLASISEDIKTLDDASVSPOTKFIILNKLNIESSIGDIYIYEKLE 844
QY 725 AFNKTISTKDQDLIKOYNTLIEEMQLVLIYVGERYPVGVNVTKEEVTREIHLCLIE 784
Db 845 PVKNIHNSIDDLIDEFNL-----ENVSELYELAKL----- 877
QY 785 PMPHSAIAKNLPE-----NENNETGLENVINKVATFKKPGVSGHGVYELKDESUKD 835
Db 878 -----NNLDEKYLISFEDISKNNSTYSVRINK-----SNGESVYVETEKEI-- 919
QY 836 FNMVYFHYSK-----TOHSAEHMQKRRKQENKDEALPPPPPFPCPAFSKV 883
Db 920 FSKYSEHITKEISTIKNSIITVDNGLNDLQDHTSQ----- 957
QY 884 INLLNCDIMMYIILTVFERAIDTD--SNLWTEGMLQMAFHILALGLLLEKQOLQKAPEE 941
Db 958 VNTLNA---AFFIQSLIDYSSNKDVLNDLSTSVKQVLAQFLSTGLNTIYDSIQLV----- 1010

QY 942 VTDFYHKASRLGSSAMN--IOMLEKLGIPQLEGOKDMITW---ILOMFDTVKRLREK 996
Db 1011 -----NLISNAVNDTINVLPITTEGPIVSTILDGNGALKELDLDEHDPDLKK 1060
QY 997 -----SCLIVATTSGSESINKNDEITHDKEAERKKAERKAEARLHROKIMQMS 1043
Db 1061 ELEAKVGVLAINNLSIAATVASIVGI-GAEVT-----IFLLPIAGISAGIP 1106
QY 1044 ALQKNFIETHKLMYDNTSEMPGKEDSIMEEESTPAVSDYSRIALGPKRGPSVTEKE-VLT 1102
Db 1107 SLVNN-----ELILHDKATSVVNYFNHLSSEKKGYPGLKTEDDKILV 1147
QY 1103 CILCOEQEVKIENNAMVLSACVQKSTALTQHRKPIELSGEALDPLFMDPDLA 1156
Db 1148 PIDDLVISEIDFNNSIKLGTG----NILAMEGGSGHTVTGN-IDHFFSSPSIS 1196

RESULT 4

US-08-405-496A-6
; Sequence 6, Application US/08405496A
; Patent No. 5919665
; GENERAL INFORMATION:
; APPLICANT: WILLIAMS, JAMES A.
; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
; TITLE OF INVENTION: NEUTROTOXIN
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/405,496A
; FILING DATE: 16-MAR-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPD-01308
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-405-496A-6

Query Match 1.7%; Score 156.5; DB 2; Length 2710;
Best Local Similarity 18.6%; Pred. No. 0.00065;
Matches 189; Conservative 154; Mismatches 354; Indels 44;

RESULT 5

US-08-915-136-6
; Sequence 6, Application US/08915136
; Patent No. 6290960

QY 288 KEDIKSHSENVSOHPLHVEV---LHSEIMAHQKFAIRLGSWMNKIMSYSSDPROIFCOAC 344
Db 355 KSEIFSKLENVNSLEIKIAFALGSI--NOALISKQSGSYTNLIVBOVKRYQFNLQH 412
QY 345 LREEPDSENPCLISRLMLWDAKLYGARKILHELIFSSPFMEWYKKLFAMEFVKYKOL 404
Db 413 LNPAIESDNN-----FTDTTKIFHDSLFNSATAE---NSMELTKIAPY---L 453
QY 405 QKEYISDDHDSRISITA-----LSVQMFTV-PTL-ARHLIE---EQNVISVITE 448
Db 454 QVGEMPEARS-TISLSGPGAYASAYDFINLQENTIEKTLKASDLIEFKFPENNLSOLTE 512
QY 449 TLLEVLPEYLDNRNKNFNQGSQDLGRVYA-----VICDLKYLILSK-PIWITE 497
Db 513 QEINSLWFDQASAKYQFEKYVRYDTGGSLSDNGVDFNKNKTALDKNYLLNNKIPSNVYE 572
QY 498 RLRMQFLEGFSLKILTCMOGMEIRQVQGHIEVDPDWEAAIAIOMLKKNILLMFQEW 557
Db 573 E-----AGSKNYVHVIIQLQDDISYEATCNLFKNP--KNSIIQRMNSESASY--F 622
QY 558 CACDEELLVAYKECHKAVMRCSTSFISSTTVQSCGH-----SLETKSYR 604
Db 623 LSDDGESIL---ELNK--YRIPERLKNKEKVKVTFIGHGKDEFNTSEFARLSVDSLSNE 676
QY 605 VSEDLVSIHLPLSRTLAGLHVRLSRIGAVSRUHEFVSPEDFOVEVLVEYPLRCLVLVAQV 664
Db 677 ISSFLDTIKLIDSPK---NVEVNLGC-----NMFSY-DFNVE--ETYPGKLLLSIMDK 724
QY 665 VA-----EMWRRNGLSLISOVFYQDVQKREEMV--- 693
Db 725 ITSTLPDVKNSITIGANOYEVRINSEGRKELLASHGKWINKEEAIMSLSKEYIFFDS 784
QY 694 -----DKDIIMLQIGASLMDPNKFLLLVLQ-----RYELAE 724
Db 785 IDNKLKAKSNIPGLASISEDIKTLILDASVDPDKFILNNLKLNESSIGYIYVEKLE 844
QY 725 AFNKTISTQDQLIKOYNLTIEMLQVLIYIVGERYVPGVGNVTEEMREIHLCLIE 784
Db 845 PVKNIHNSIDDLIDFNLL-----ENVSDLEYELKKL----- 877
QY 785 PMPHSAIAKNLPE-----NENNETGLENNVINKVATFKPGVSGHGYELKDBSLKD 835
Db 878 -----NNLDEKYLISFEDISKNNSTYSVRFINK-----SNGESVYVETEKEI-- 919
QY 836 FMNIEYHYSK-----TQHSKAEHMOKKRRKQENKDEALPPPPPEFCPAFSKV 883
Db 920 PSKYSEHITKEISTIKNSIITDNGNLLDNIQLDHTSQ----- 957
QY 884 INLNCIDIMMYILRTVFERAIDTD--SNLWTEGMLQMAFHIALGLLEBKQQLQKAPEE 941
Db 958 VNTLNA---AFFIQSLIDYSSNKNVDLNLSTSVKQVLAQLFSTGLNTIYDSIQLV--- 1010
QY 942 VTDFYHKASRLGSSAMN--IOMLEKLGIPQLEGOKDMITW---ILOMFDTVKRLREK 996
Db 1011 -----NLISNAVNDTINVLPITTEGPIVSTILDGNGALKELDLDEHDPDLKK 1060
QY 997 -----SCLIVATTSGSESINKNDEITHDKEAERKKAERKAEARLHROKIMQMS 1043
Db 1061 ELEAKVGVLAINNLSIAATVASIVGI-GAEVT-----IFLLPIAGISAGIP 1106
QY 1044 ALQKNFIETHKLMYDNTSEMPGKEDSIMEEESTPAVSDYSRIALGPKRGPSVTEKE-VLT 1102
Db 1107 SLVNN-----ELILHDKATSVVNYFNHLSSEKKGYPGLKTEDDKILV 1147
QY 1103 CILCOEQEVKIENNAMVLSACVQKSTALTQHRKPIELSGEALDPLFMDPDLA 1156
Db 1148 PIDDLVISEIDFNNSIKLGTG----NILAMEGGSGHTVTGN-IDHFFSSPSIS 1196

GENERAL INFORMATION:
APPLICANT: KINK, JOHN A.
APPLICANT: THALLEY, BRUCE S.
APPLICANT: PADHYE, NISHA V.
APPLICANT: FIRCA, JOSEPH R.
APPLICANT: STAFFORD, DOUGLAS C.
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND PREVENTION OF C. DIFFICILE DISEASE
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/915,136
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/480,604
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01763
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-915-136-6

Query Match 1.7%; Score 156.5; DB 3; Length 2710;
Best Local Similarity 18.6%; Pred. No. 0.00065;
Matches 189; Conservative 154; Mismatches 354; Indels 317; Gaps 44;
QY 288 KEDIKSHSNVQHPHLEBV---LHSEIMAHQKFAIRLGLGSMWNKIMSYSSDFRQIFCQAC 344
DB 355 KSEIFSKLENLVNLSLEIKIAGLSVI--NQALISKQGSYTLNVLIEQVKRYQFLNOH 412
QY 345 LREEDPSNCPILSRMLMDAKLYGARKILHELIFSSPFMEYKKLFAMEFVKYKQL 404
DB 413 LNPAIESDNN-----FDTTKIFHDSLFNSATAE--NSMELTKIAPY---L 453
QY 405 QKEYISDDHRSISITA-----LSVQMFTV-PPL-ARHLIE---EQNVISVITE 448
DB 454 QVGFMPEARS-TISLSGCGAVASAYDFINLQENTIEKTLKASDLIEFKFPENNLSQLTE 512

QY 449 TLLEVLPEYLDNRNKNFQYSDQKLGKRVYA-----VICDLKYLISK-PTIWIWE 497
DB 513 QEINSLWSFDQASAKYQFEKYVRDYTGSGLSSENGVDNFKNTALDKNYLLNKNKIPSNVVE 572
QY 498 RLRMQFLEGFRSFLKILTCMQGMEIEIRQVQGHLEVPDWEAAIAIOMOLKNILLMFQEW 557
DB 573 E-----AGSKNVVHYIIQOGDDISYEATCNLFESKNP--KNSIIQRNMNESAKSY--F 622
QY 558 CACDEELLVAYKECHKAVMRCSTSFISSTKTVVQSCGH-----SLETYSYR 604
DB 623 LSDDGESIL-----ELNK--YRIPERLKNKEKVKVTFICHGKDEFTSEFARLSVDSLSNE 676
QY 605 VSEDLVSIHPLSRTLGLHVRLSRIGCAVSRHLFEVSEFEDQVEVLYEYPLRCLVLVAQV 664
DB 677 ISSFLDTIKLDISPK-----NVEVNLGCG-----NMFYSY-DFNVE--ETYPCKLLLSIMDK 724
QY 665 VA-----EMRRNGLSLISQVFFYQDVQKCREMY--- 693
DB 725 ITSTLPDVKNKSITIGANQVEYRINSEGRKELLASHGKWINKKEAINDLSKSEYIFPDS 784
QY 694 -----DKDITMLQIGASLMDPNKFLLLVLQ-----RYELAE 724
DB 785 IDNKLKAKSNIPGLASISEDIKTLLDASVSPDTKFLNKLNISSIGDYIYERKLE 844
QY 725 AFNKTISTKDQDLIKQYNTLIEEMQLVLIYIGERYVPGVGNVTKEEYVMEIHLICIE 784
DB 845 PVKNIHNSIDDLIDEENLL-----ENVSDELYELKLL----- 877
QY 785 PMPHSAIAKNLPE-----NENNETGLENVINKVATPKPGVSGHGVYELKDESLKD 835
DB 878 -----NNLDEKYLISFEDISKNNSTYSVRFINK-----SNGESVYVETEKEI-- 919
QY 836 FNMVYFHYSK-----TOHSAEHEMKRRKQENKDEALPPPPPEFCFPAFSKV 883
DB 920 FSKYSEHITKEISTIKNSIITDVNGNLLDNLQDLHTSQ----- 957
QY 884 INLLCNDIMMYILRTVFERAIDTD--SNLWTEGMLQMAFHILALGLLEEKQOLOKAPEEE 941
DB 958 VNTLNA---AFFIQSLIDYSSNKDVLNDLSVQVLAQVLFSTGLNTIYDSIOLV----- 1010
QY 942 VTDFYHKASRLGSSAMN--IQMLEKLGIPLEGQKDMITW---ILOMEDTVYKRLREK 996
DB 1011 -----NLISNAVNDTINVLPITTEGIPVSTILDGINLGAATKELLDEHDPPLKK 1060
QY 997 -----SCLIVATTSGSEIKNDEITHDKEKAEKRAEAAARLHRQKIMQMS 1043
DB 1061 ELEAKVGVLAJNMSLSIAATVASIVGI-GAEVT-----IFLLPIAGISAGIP 1106
QY 1044 ALQKNFIETHKLMYDNTSEMPGKEDSIMEESTPAVSYSRIALGPKRGPSVTEKE-VLT 1102
DB 1107 SLVNN-----ELILHDKATSVNVFNHLSSEKSKYGLKTEDDKILV 1147
QY 1103 CILCQEEQEVKNIENNAWVLSACVQKSTALTQHRKPIELSGEALDPLFMDPDLA 1156
DB 1148 PIDDLVISEIDFNNSIKLCTC-----NILAMEGGSGHTVTGN-IDHFFSPSPSIS 1196

RESULT 6

US-08-957-310-6
Sequence 6, Application US/08957310
Patent No. 6365158
GENERAL INFORMATION:
APPLICANT: Williams, James A.
APPLICANT: Kink, John A.
TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES
TITLE OF INVENTION: OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE
TITLE OF INVENTION: DISEASE
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medlen & Carroll
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco

STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/957,310
FILING DATE: 23-OCT-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 24-OCT-1994
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01121
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-957-310-6

Query Match 1.7%; Score 156.5; DB 4; Length 2710;
Best Local Similarity 18.6%; Pred. No. 0.00065;
Matches 189; Conservative 154; Mismatches 354; Indels 317; Gaps 44;
Qy 288 KEDIKSHENVSOHPLHVEV---LHSEIMAHQKFAIRLGSWMNKIMSYSSDFRQIFQAC 344
Db 355 KSEIFSKLENLVSDLEIKIAFALGSLVI--NQALISKQSGYLTNIVIEQVKNRYQFLNQH 412
Qy 345 LREPDSNPCLISRLMDAKLYGARKILHELIFSPFFMEMEYKFLFAMEFYKYLQ 404
Db 413 LNPAGESDNN-----FTDTTKIFHDSLFNSATAE---NSMFLTKIAPY---L 453
Qy 405 QKEYISDDHRSISITA-----LSVOMFTV-PTL-ARHLIE---EQNVISVITE 448
Db 454 QVGFMPERS--TISLSGPGAYASAYDFINQENTIEKTLKASDLIEFKFPENNLSQUTE 512
Qy 449 TLEVLPEYLDNRNKNFQGSQDKLGRVYA-----VICDLYILISK-PTIWE 497
Db 513 QEINLSWSDQASAKYQEKVRYDTGGSLEDNGVDENKNTALDKYLLNNKIPSNVE 572
Qy 498 RLRMQFLEGFRSFKILTCMQGMEIRKQVGHIEVDPDPAATIAQMLKNILMFQEW 557
Db 573 E-----AGSKNVHYIIQLQDDISYEATCNLFKSNP--KNSIIQRNMNESAKSY--F 622
Qy 558 CACDEELLVAYKECHRAVMCRSTSFSSSTTVVQSCGH-----SLETKSYR 604
Db 623 LSDDGESIL-----ELNK--YIPRLKNKKEVKVTFIGHGKDEFNTSEFARLSVDLSNE 676
Qy 605 VSEDLVSTHPLSLRTLACGLVRLSLGAVSRHLHFVSPEDFQEVVLVEYPLRCLVLVAQV 664
Db 677 ISSFLDTIKLDISPK-----NVEVNLGC-----NMFSY-DFNVE---ETYPKLLLSIMDK 724
Qy 665 VA-----EMWRNRGLSLISQVFFYQDVKRCREMY--- 693
Db 725 ITSTLPDVNKNISITIGANOYEVIRINSEGRKELLAHSGKWINKEAIMSDLSKEYIFFDS 784

Qy 694 -----DKDIIMLIQIASLMDPNKFLLLVLO-----RYELAE 724
Db 785 IDNKLKAKSKNIPGLASISEDIKTLTLLDASVSPDTKFIILNNKLNTSESSIGDYIYYEKLE 844
Qy 725 AFNKTISTKDDOLIKOYNFLIEEMQLVLIYIGERYVPGVGNVTKREVTMRREIHLCLIE 784
Db 845 PVKNIHNSIDDLIDFENLL-----ENVSDELYELKKL----- 877
Qy 785 PMPHSAIAKNLPE-----NENNETGLENNVINKVATFKRPGVSGHGVYELKDESLKD 835
Db 878 -----NNLDEKYLISFEDISKNNSTYSVRFINK-----SNGESVYVETEKEI-- 919
Qy 836 FNMVYFYHSK-----TOHSAEAMHOKKRRKQENKDEALPPPPPEFCPAFSKV 883
Db 920 FSKYSEHITKEISTIKNSIITDVNGNLLDNIQLDHTSQ----- 957
Qy 884 INLLNCDIMMYILRTVFERAIDTD--SNLWTEGMLQMAFHIALGALLEKQOQLOKAPEE 941
Db 958 VNTLNA--AFFIQSLIDYSSNKNLDNLSTSVKQVLYAOLFSTGLNTIYDSIQLV---- 1010
Qy 942 VTFDFYHKASRLGSSAMN--IQMLLEKLGIPQLEGQKDMITW---ILOMFDTVKRLREK 996
Db 1011 -----NLISNAVNDTINVLPITIEGPIVSTILDINGLGAIAKELLDEHDPPLKK 1060
Qy 997 -----SCLIVATTSGSESINKDEITHDKEAKERRKABAAARLHRQKINMAOMS 1043
Db 1061 ELEAKVGVLAINNLSLSIAATVASIVGI-CAEVT-----IFLLPIAGISAGIP 1106
Qy 1044 ALQKNFIETHKLMYDNTSEMPGKEDSIMEEESTPAVSYSRIALGPKRGPSTVEKEP-VLT 1102
Db 1107 SLVNN-----ELILHDKATSVVNYFNHLSSEKKGYPGLKTEDDKILV 1147
Qy 1103 CILCOEEQVKIENNAMVLSACVQKSTALTOHRGKPIELSGEALDPLFMDPDOLA 1156
Db 1148 PIDDLVISEIDFNNSIKLGTC-----NILAMEGGSGHTVTGN-IDHFFSSPSIS 1196

RESULT 7

US-10-011-366-6
; Sequence 6, Application US/10011366
; Patent No. 6573003
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Kink, John A.
; TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES
; OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE
; DISEASE
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/011,366
; FILING DATE: 16-NO. 6573003-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/957,310
; FILING DATE: 23-OCT-1997
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 24-OCT-1994
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; APPLICATION NUMBER: US 07/985,321

Db 725 QMSALQAKLDEEBHKNKLOQHVD-----KLEHHSQMQELFSSERIDWTQK 771
QY 561 DEELL---VAYEKCHAVNRSTSPISSTKVVQSCGHS-----LETSKYRVSED 608
Db 772 QEBLLSQLNVLEKQLOET--QTKNDFLKSEVHDLRVVLHSAKELSSVKLEYSFSTQOE 829
QY 609 -----LVSIHLPLSRTLGLHVLRLSRL-----GAVSRLEHFVSFE-----DFQ 646
Db 830 KEFNKLSERHMVQLQDLNLRLENKLESKACQSDYDNLQELFMKEFIDQLSRNLQNF 889
QY 647 VE-VLVEYPLRCLVLVAQVAVEMWRRNGLSL-----ISOVFYQDV 686
Db 890 KENETLKSDNLNLMELLE--AEKERNKLSLQFEEDKENSKEILLKVLAVRQEKQETA 947
QY 687 KCREM-----YDKDIIMLQIGASLMDPNKFL--LVLRQYELAEAF---NKT 729
Db 948 KCEQMAKVKLEESLATEKVISSLE---KSRDSKKVVADLMNQIQELRSSVCEKTET 1004
QY 730 ISTKQD---DLIKQYNTLI--EEMQLVLIYIGERYVPGVGNVTKEV---TMREIHL- 780
Db 1005 IDTLKQELKDINCKYNSALVDREESRVLI-----KKQEVLDLKLKTLRL 1050
QY 781 -----LCIEPMPHSAIAKN-LPENENNETGL-----ENVINKVATEFKPGVSGHG 824
Db 1051 ILSEDIERDMLC-EDLAHAQELNMLTEASKHSLQSAQBELTKKEALIQE---LQHK 1106
QY 825 VYELKDESUKDFNMFYHYSKTOH---SKAEHMOKRRKQENKDEALPPPPPPFCPAPS 881
Db 1107 LNOKKEVEQKKNYFNKMRQLEHVMSAAEDPOS-----PKTPPHFQTHLA 1153
QY 882 KVINLLNCDIM-MYILRTVFERAI---DTDSNLWTEGMLQMAFHILALGLLEEQLOKA 937
Db 1154 KLETOEQEIEDGRASKTSLHLVTKLNEDEVKNAILRM-----KEQLE- 1200
QY 938 PEEVTFDFYHKASRLGSSAMNIQMLLEKLGIPQLEGQKDMITWILQ-MEDTVKRLREK 996
Db 1201 -----MENLRLESQOLIEK-----NWLQGOQLDDIKRQKE- 1230
QY 997 SCLIVATTSSESINKDEITHDKAEKRRKAEAAARLHROKIMAQMSALQKNFTHTKLM 1056
Db 1231 -----NSDONHPDNOQLKNEQESIKE-----RLAKSKIVEMLKMKADLEEVQSAL 1277

RESULT 10

US-09-723-262-2

; Sequence 2, Application US/09723262
; Patent No. 6379912
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: No. 6379912el motor proteins and methods for
; FILE OF INVENTION: their use
; FILE REFERENCE: 1017
; CURRENT APPLICATION NUMBER: US/09/723,262
; CURRENT FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: US 09/572,191
; PRIOR FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1388
; TYPE: PRT
; ORGANISM: Human
US-09-723-262-2

Query Match 1.6%, Score 150.5, DB 4; Length 1388;
Best Local Similarity 19.3%, Pred. NO. 0.00073;
Matches 212; Conservative 162; Mismatches 372; Indels 355; Gaps 54;
QY 187 LNEEVIQARKIPSVIKYVVMETIWEEE--KE-----LPELOIREKNEYCYVLFND- 238
Db 447 LKREKFIQSNKM---IVKRFEDQIIRLEKLHKSRRGGFLP-----EQDRLSRLREI 497
QY 239 -----EHSYDHVIYSLQALDCELAEAQHTTAIDKGRRAVAKAGAYAACQAKEDI 291
Db 498 QTLREQLEIHP-----RVAKYAMENHSLREENRL-----RLLEPV 533
QY 292 KSHSENVSQHPLHVEVLHSEIMAHQKFAIRLGWSMNKIMSYSSDFRQIFQACALREEDPS 351
Db 534 KRAQEMDAQTIAKLEKAFSEISGMEK-----SDKNOQGFSPKAQKE----- 574
QY 352 ENCLLSRLMDAKLYGARKILHELIFSSFFMEMEYKKLF--AMFVYIYKQLOKEY 408
Db 575 --PCLFANTEKLAQALL---QIOTELNNSKQYE-EFKELTRKROLESELOSLQAN 627
QY 409 IS-----DDHRSISITALSVMQFTVPTLARHLIEBQNVISVITELLEVLP 455
Db 628 LNLNLEATKAKRQEVSQLNKLHAETKLIITPTTKAYQL-----HSRPVPKLSP 678
QY 456 EYLDNNKNFQGYSDQKLGVRV---AVICDLKYILISKPTIWTLELRMOFLEGRFSFLK 512
Db 679 E-----MGSGSLYTONSILLNDILNEPVPPEMNEQAFEAISEELRTVQE 724
QY 513 ILTCMQGM-----PEIRROVQOHIEVDPWEAAIAIQOLKNILLMFQ-----WCAC 560
Db 725 QMSALQAKLDEEBHKNKLOQHVD-----KLEHHSQMQELFSSERIDWTQK 771
QY 561 DEELL---VAYEKCHAVNRSTSPISSTKVVQSCGHS-----LETSKYRVSED 608
Db 772 QEBLLSQLNVLEKQLOET--QTKNDFLKSEVHDLRVVLHSAKELSSVKLEYSFSTQOE 829
QY 609 -----LVSIHLPLSRTLGLHVLRLSRL-----GAVSRLEHFVSFE-----DFQ 646
Db 830 KEFNKLSERHMVQLQDLNLRLENKLESKACQSDYDNLQELFMKEFIDQLSRNLQNF 889
QY 647 VE-VLVEYPLRCLVLVAQVAVEMWRRNGLSL-----ISOVFYQDV 686
Db 890 KENETLKSDNLNLMELLE--AEKERNKLSLQFEEDKENSKEILLKVLAVRQEKQETA 947
QY 687 KCREM-----YDKDIIMLQIGASLMDPNKFL--LVLRQYELAEAF---NKT 729
Db 948 KCEQMAKVKLEESLATEKVISSLE---KSRDSKKVVADLMNQIQELRSSVCEKTET 1004
QY 730 ISTKQD---DLIKQYNTLI--EEMQLVLIYIGERYVPGVGNVTKEV---TMREIHL- 780
Db 1005 IDTLKQELKDINCKYNSALVDREESRVLI-----KKQEVLDLKLKTLRL 1050
QY 781 -----LCIEPMPHSAIAKN-LPENENNETGL-----ENVINKVATEFKPGVSGHG 824
Db 1051 ILSEDIERDMLC-EDLAHAQELNMLTEASKHSLQSAQBELTKKEALIQE---LQHK 1106
QY 825 VYELKDESUKDFNMFYHYSKTOH---SKAEHMOKRRKQENKDEALPPPPPPFCPAPS 881
Db 1107 LNOKKEVEQKKNYFNKMRQLEHVMSAAEDPOS-----PKTPPHFQTHLA 1153
QY 882 KVINLLNCDIM-MYILRTVFERAI---DTDSNLWTEGMLQMAFHILALGLLEEQLOKA 937
Db 1154 KLETOEQEIEDGRASKTSLHLVTKLNEDEVKNAILRM-----KEQLE- 1200
QY 938 PEEVTFDFYHKASRLGSSAMNIQMLLEKLGIPQLEGQKDMITWILQ-MEDTVKRLREK 996
Db 1201 -----MENLRLESQOLIEK-----NWLQGOQLDDIKRQKE- 1230
QY 997 SCLIVATTSSESINKDEITHDKAEKRRKAEAAARLHROKIMAQMSALQKNFTHTKLM 1056
Db 1231 -----NSDONHPDNOQLKNEQESIKE-----RLAKSKIVEMLKMKADLEEVQSAL 1277

; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: HUMAN
; US-08-353-700-1

Query Match 1.6%; Score 147.5; DB 1; Length 3248;
Best Local Similarity 18.3%; Pred. No. 0.0058;
Matches 318; Conservative 253; Mismatches 632; Indels 537; Gaps 81;

QY 171 HEPGRAGTIKNSRCPNLEE---VIVQARKIFPPSVIKYVEMTWEEE---KELPELQ 223
DB 304 HEKEMKGVNKTQELQLEKAKVELIEKEVLNKCRDELVRTTAQYDQASTKYTALEQK 363
QY 224 TREKNERYCVLFNDE--HHSYDHYIYSLQALDCELAQAHLHTTAIDKEGRAVAKAGAY 281
DB 364 LKKLTEDLSCORQNAESARCSLEQKIKEKEKEFEELSQRQSFQTLDOECIQ-MKARLT 422
QY 282 AACQAKEDIKSHSENVSOPHLVHVEIWAHQFALRLGSMWNKIMSYSDRQIFC 341
DB 423 QELOQAKN-----MH-NVLQAEID-----KLTSVKQOLENNLEEFKQKLC 461
QY 342 QACLRPEPDPENPCILSLRLMLDALKYLGARKILHELIFSSPFMEYKKLFAFEVKKY 401
DB 462 RA-----EQAFQASQIKENELRR--SMEEMKE 487
QY 402 KOLQREYISDDHRSISITALSQVMTVP---TLARHLIEBQNVISVITETLEVLPEYL 458
DB 488 NNLLK---SHSQKAREVCHLAELKNIKQCLNOSQNFAPAEKAKVTSQETMLRDQEE-- 542
QY 459 DRNNKFNQGYSDQKLGRRVAVICDLYILISKPTIWTERRLMQFLEGRSFILKILTCMQ 518
DB 543 ----KINQOENSL-TLEKLLKAVADLE-----KQRDCSQ 571
QY 519 GMEELRRQVGHIE-----VDPDWEAAI-AIQM-----QLKNILLMFQWCADEE 563
DB 572 DULKKRE---HHIEQLNDKLSKTEKESKALLSALELKKEYBELKEETLFCWCKSENEK 628
QY 564 LLLVAYKECHKAVMRCSTSFSSKTVQSCGHSLETKSYRVEDLSVSIHLP---LSRTL 620
DB 629 LL-----TQMESEKENLQSKINHLETCLTKTOIKSHYENRVPTLEMDRENLSVEI 679
QY 621 AGLH-----VRLSRILGAVSRHLHEFVSFEDQVEVLVE-YPLRCLVLVAQVVAEMWR 670
DB 680 RNLHNVLDSSKSEVEVETQKL-AYMELQKAEFSQKQKQKEIENMCLNKLTSQITQGVED--L 735
QY 671 RNLGLSISOVFFYQVQKREEMVD-----KDIIMLIQIGASLM---DPNKLFLVLVQRYEL 722
DB 736 EKLQDLNEIMDKD-RCYQDLHAEVESRLDLKSK-DASLVNEDHQRLSLLAFOQOPAM 793
QY 723 AEAFNKTI-----STKDQDLI-----KOYNTLIEEML 749
DB 794 HHSFANIIGEQQSMPSERSECRLEADQSPKNAILQNRVDSLEFSLQSKQMNSDIQKQC 853
QY 750 QVLIYIVGGRYPVGVGNVTKEEVTMRITHLICI-----EPMHP-SAIANKLP 796
DB 854 EELVOLKGE-----IEENLMKAEQHQSFVAETSORISKLOEDTSAHQNVVAETLS 904
QY 797 ENNETNETGLYENKIVAT-----FKKPGCVSGHYELKDESLKDFNMVFFYHYSKQHS 849
DB 905 ALENKEKELQDLNDKVEVETQAEIQLKK---SNH-----LLEDLSKELQLL-----S 948
QY 850 KAEHQKRRKQENKDEALPPPPPPFCFAPAFSKVINLNCIDIMMYILRTVTFERAID--TD 907
DB 949 ETLSLSEKKE-----MSSIISL-----NKREIEELTQ 974

QY 908 SNLWTEGMLQWAFHILALGLLEEKQOQLOKAPBEEVTFDFYHKASRLGSSAMNIQMLLEKL 967
DB 975 EN-----GTLKE-----INASLQERKMNIQKSE---SFANYIDEREKSISELSQYKQEK 1023
QY 968 KGIPQEGQKDMITWILQMFDTVKRLREK-SCLIVATTSGSESIGND---EITDHKEKA 1022
DB 1024 ILQRCETGNAYEDLSQYKAAQEKNSKLECLNNECTSLCENRKNRNEQLKEAKFAHQ 1083
QY 1023 ERKKAABAARLHRQKIMASALQKNFTIETHKLMYDNTSEMPGKEDSIMEESTPAVS 1082
DB 1084 EFLTKLAFABERNQNLMELETVOQ---ALRSEMTDNQNNKSEAGGLKQEIEMT----- 1134
QY 1083 SRTALGPKRGPSTVEKEVLTICLQEBQEVKIENNAVLSACVQKSTALTQHRGKPIELS 1142
DB 1135 -----LKEEQNKQKQEV-NDLQENELQMKVMK-----TKHE----- 1165
QY 1143 GEALDPLFMDPLAYGTYTGSCGHVMAVQKFEAVOLSSQQRHIVDLDFLEGEVLC 1202
DB 1166 -----CONLESEPIRNSVKER-----ESE----- 1184
QY 1203 PLCKSLCNTVPIIPILOPKQKINSNADA-LAQLLTLLAR-----WIQT 1243
DB 1185 ---RNQCN-FKPQMDLEVKEISLDSYNAQLVQLEAMLRNKLKQSEKECEQLQHELQT 1240
QY 1244 VLARISYNTIRHAKGNPIPIFPNOGMDSTLEFHSILSFGVSESSIKYSNIKEMVILFA 1303
DB 1241 IRGDLSTSLNDCMQSQE-----ISGLKDCIDAEERYISGPHELSTSONDAHLQCSIQ 1294
QY 1304 TTIYRIGLVKPPDERDPRVPMLTWSTCAFTIQAENLGLDGEKPLFGALQNRHQLKAL 1363
DB 1295 TTMNKLN-----ELEKICEILOAEKYEVLTELNDRSRCITAT 1332
QY 1364 MQFAVAQRITCPQVLIQKHLVRLSVLPNLTSEPTCLLSIDLPHVLVGAFLAPSLYW 1423
DB 1333 RKMA-----EEVGKLLNEV--KILNDDSG-LLHCELVEDIPGGEFG-EQPN 1375
QY 1424 DDPVDLPSSVSSYNHLYLPHLITMAHMLQIILTVDTGLPLAQVQDESEEAH-SASFF 1482
DB 1376 QHPVSLAPLDESNEYHLTSDKEVQMFAEL---QEFSLSLQSEHKILHDQCMSSKM 1432
QY 1483 AEISOYTSG-----STGCCDIPGWY-----LWVSLKNGITPYLRCAAL----- 1519
DB 1433 SELQTYVDSKAEANLVSTNLRNFQGLVEMQGLEGLVPSLSSSCVPDSSSSLSLGD 1492
QY 1520 --FFHYLLGVTPPELHTNSAEGYSA-----LCSYLS-----LPTNLF 1557
DB 1493 SSFYRALLEQTGDMSSLN-LEGAVSANQCSVDEVCSSSLQTYVDSLKAENLVSTNL-R 1550
QY 1558 LFOEYWDTVR-----PLIQRCA-DPALLNCLKQKNTVVYPRKRNLSIELPDD 1605
DB 1551 NFQ-GDLVKEMQLGLEGLVPSLSSSCVPDSSSSLSLGDSSFY-----RALLEQTGD 1601
QY 1606 YSCLLNOASHRCPRSDDERKHPVLCFLFCAILCSQNICQOEIVNGBEVGACIFALHLC 1665
DB 1602 MSLLSNLEGVYVANSQCSVDE-----VFCSS-LQENLTRKETSPAKGVVEELES-- 1651
QY 1666 KARGCAYPAPVDEYGE--TDPGLKRGNPLH-----LSRER-----VMOQ 1709
DB 1652 ----CEVYROSLEKLEKEMESQGMKNKEIOELQLLSERQELDCLRKQYLSENEQWQ 1707

RESULT 13

PCT-US95-16216-1
; Sequence 1, Application PC/TUS9516216
; GENERAL INFORMATION:
; APPLICANT: Yen, Timothy J.
; APPLICANT: Rattner, Jerome B.
; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
; TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720

Db 1602 MLLSLNLEGVWSANQCSVDE-----VFCSS-LOENLTKRETPSAPAKGVLEESL-- 1651
QY 1666 KARGCAYPAPYDEYGE--TDGLKRGNDPLH-----LSRER-----YRKLHL-----VWQQ 1709
Db 1652 ----CEVYRQSLEKLEKMEGOGIMKNKEIQLEQLLSRQELDLCKLRQYLSENEQWQQ 1707
RESULT 14
US-09-914-259-67
; Sequence 67, Application US/09914259
; Patent No. 6495336
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/09/914,259
; CURRENT FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 67
; LENGTH: 959
; TYPE: PRT
; ORGANISM: Rattus norvegicus
; US-09-914-259-67

Query Match 1.5%; Score 140.5; DB 4; Length 959;
Best Local Similarity 20.6%; Pred. No. 0.0031;
Matches 155; Conservative 112; Mismatches 256; Indels 231; Gaps 37;
QY 526 QVGOHIEVDPPWEAIAIQMOLKNILLMFQEWACADEILLV-----AYKECHKAV 576
Db 264 EVG---DENPGWSA-----QKVTNLHLMQ-----LVRLVSPTPPPGATSSCQKAM 307
QY 577 MRCS-----TSFSSKTVVQSGHSLFTKRYVSED-LVSIHLPL 616
Db 308 FQCGLLQQLCTILMATGIPADILTTINTVSEVIRG-----QVNQDYFASVNA 358
QY 617 SRTLAGLHVRSLRGAVSRLEHFFSVFQEVVLPCLVLAQVVAEMWRNGLSL 676
Db 359 NPPRAIVVLL--MSWNERPPFV-----LRCAVLYC----- 388
QY 677 ISOVFYODVCKREMYDKDIIMLQIGAS-----LMDPNKELLVLQRY-----ELAE 724
Db 389 -FQCLYKNEKGQGEIV-ATLLPSTIDATGNSVSAGQLCGGLFSTDSLSNCAAVALAH 446
QY 725 AFNKTISTKDDQLIKQYNTLI-----EEMQLVLIYIV--GERYVPGVGN-----VTKEE 771
Db 447 ALOGNATQKEQLLRVOLATSIGNPPVSLLOQCTNLSQGSKTQTRVGLLMLLCTWLSNCP 506
QY 772 VTMREIIHLLCIEPMPHSAIAKNLPENNETGLENVINKVATFKKPGVSGHGVYELKDE 831
Db 507 IAVTHFLHNSANVPPLTQGAENLGEELQVGLCALL-----LGTSIY-FNDN 554
QY 832 SLKDFNMFYHYSKTQHSKAEHMOKRRKQENKDEAL-----PPP-PPP 874
Db 555 SLE-----NYTK---EKLQOLIEKRIGKENYIEKLGFIKSHKEYLSRASQRPQPNFPSP 604
QY 875 E---FCPAFSKVINLNCDDIMYILRTVFE-----RAITDSNLTMEGLMAFHIL 923
Db 605 EYMFDPHEFTKLKVELEGVITKAIYKSSSEEDKKEEVKKTLEQHNIVTHYKNMTRQDL 664
QY 924 ALGLEEK-----QOLQKAPEEVT-----FDFYHKASRLGSSAM 958
Db 665 QLEELKQVSTLKQNEQLOQTAVTQOASIQQHKDQYNLLKVQLGKDNHGHQSHSDGAQV 724
QY 959 N-----TOMLEKLUKGPQEGQKDMTWILQNFDTVKRLREKSLCIIVATTSGSESIGN 1012
Db 725 NGIQPEEISRLREE---IEELRSHQVLLQ-----SOLAEKDTVI-----ENLRS 765
QY 1013 DEITHDKAEK---RKRKAEARLHRQIMQMSALQKNFIETHKLMDYNTSEMPGCKDS 1069

Db 766 SQVSGMSEQALATWCSPRDAQVAVELKQELSALKSOLCSQSLEITRLQTEN-SEIQORDET 824
QY 1070 IMBEEESTPAYSDYSRIALGPKRGPVSVTEKEVLTILCQEEQEVKIENNAMVLSACVQKST 1129
Db 825 L--AKSVP-VEGESLVTAAK-----TTDVEGRLSALLQETKELKNE-----IKALSEERT 872
QY 1130 ALTQHRGKPIELSGEALDPLFMDPDLAYGTYTGS 1163
Db 873 AIQ-----KQLDSSNSTIAILOTEKDKLYLEVTD 902
RESULT 15
US-08-493-092-2
; Sequence 2, Application US/08493092
; Patent No. 5728807
; GENERAL INFORMATION:
; APPLICANT: Shiloh, Yosef
; APPLICANT: Tagle, Danilo A.
; APPLICANT: Collins, Francis S.
; TITLE OF INVENTION: Ataxia-Telanglectasia Gene
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Reising, Ethington, Barnard & Perry
; STREET: P.O. Box 4390
; CITY: Troy
; STATE: Michigan
; COUNTRY: US
; ZIP: 48099
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/493,092
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kohn, Kenneth I.
; REGISTRATION NUMBER: 30,955
; REFERENCE/DOCKET NUMBER: P-310 (TAV)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810) 689-3500
; TELEFAX: (810) 689-4071
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1708 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-493-092-2

Query Match 1.5%; Score 140.5; DB 1; Length 1708;
Best Local Similarity 17.9%; Pred. No. 0.0083;
Matches 260; Conservative 197; Mismatches 469; Indels 529; Gaps 68;
QY 324 SWNKMYSVSSDFRQIFCQACLRB-----EPDSENP-----CL-----ISRLMLWD 364
Db 509 SWRNLLSTHV-----QGFFTSCLRHFSQTSRSTTTPANLDSSEHFFRCCLDKKSQRTMLAV 564
QY 365 AKLYKGARKILHELIFS-SFFMEMEYKKL-----FAMEFVYKVKQLOKEYISDD 412
Db 565 VDMRQKRQSPSSSTINDAFWLDLNYLEVAKVAQSCAAHFTALLYAEIY---ADKKSMDDQ 622
QY 413 HDRSISITALSVMQFTVPTLARHLIEQNVISVITETLLLEVLPEYLDNRNNKFNFGYSQD 472
Db 623 EKSLAFEEGS-QSTISSISEKSEETGI---SLQDILLEIYRIGEPDSLYGCGG----- 675
QY 473 KLGRRVAVICDLKYLILSKPTIWTFLRMQFLEGFSFLKTLTCMQGMEETRRQVGQ-HI 531
Db 676 --CKMLQPTITRLR--TYEHEAMWKGALVTYDLE---TAIPSSRQAGTIQALQNLGLCHI 728

Qy	532	EVPDPWEAAIAIQOLKNIILMFQEWCADELLVA-----YKECHKAVMRC-STSF	58A
Db	729	-----LSVYLGGLDYENKDMCPLEELHLYQAARWQMWDHCTSVSKEVEGFSYH	777
Qy	585	SSSKTVVQSCG-----HSLETKSVYSEDIVSLPLSRLAGLHVLSRLGAVSRL	636
Db	778	ESLYNALQSURDREFSTFYESLKYARKEYEEMCKRSL-----ESVSYLPTLSRLQAIGEL	834
Qy	637	HEFVSPFEDFOV-----EVLVEYPLRCLVLVAQVVAEMWRRNGLSLISQVYQDV---	686
Db	835	ESTIGELFSRSVTHRQLESEVIK-----WQKHQOLLKDSFSEQEPIMAL	878
Qy	687	-----KCREEMYDKDIIMQIGASLMDPNKFLLLVLQRYELAEAFNK	728
Db	879	RTVILEITLMEKEMDNSORECIKDILTKHLVELSI-----LARTFKN	919
Qy	729	T-ISTRDQDILLKOYNTL-----TEEMLOVLIYIVGERYPGVGNVTKEEVTMRILH	779
Db	920	TQUPERAIFQIKOYNSVSCGVSEWQLEE-AQVF-----WAKKEOSLASLIL	964
Qy	780	LLICIEPMPHSAIAKN-----LPENENNETGLENVINKVATFFKPGVSGHGVY-	826
Db	965	KQMIKKLDASCANNPSLKITYECLRVCGNWLAETCLENPVIMQTYILEKAVEVAGNYD	1024
Qy	827	-----ELKDSLKDFNMFYHYSKTQHSKAEHMOK-----KRRKQE-----	862
Db	1025	GESSDELNRNGKMAF-LSLARFSDTOYORTENYMKSEFENKQALLKRAKEVEGLLREHK	1083
Qy	863	-----NKDE-----ALPPPPPEFCAPSKVINLNLCDIM-----MYILRTV	899
Db	1084	IQTNRYTVKQORELEDELALRALKDEKRFCKA-----VENYINCLLSGEEHDMWVFL-	1139
Qy	900	FERAIDTDSNLWTEGMLQMAFHILALGLLEEKQOLOKAPEEVTFDP-----YHKASPLGS	955
Db	1140	-----CSLWLENS-----GVSFVNGMKMRGDKMIPTYKPLMLYQLAARNGT	1181
Qy	956	SAMNIQMLEKLKIGIPOLEGOKDMITWLOMFDTVKRLREKSLDIVATTSGSESIKNDEI	1015
Db	1182	KMWG-----GLGFHEVLNLLISRI-----SMDHPHHTLFIILALANAN---RDEF	1223
Qy	1016	THDKEKAERK-----KAEAA-----RLHROKIMQMSALOKNFI---1050	
Db	1224	LTRPEVARRSRITKYNVPKQSQSLDEDETEAANRIICTIRSRPQMSRVSEALCDAYILA	1283
Qy	1051	-----ETHKLMVDNTSEMPGKEDSTIMEESTPAV-----SDYS-----R	1084
Db	1284	NLDATOWKTORKGINPADQPIYTKLNLIEDVVVPTWEIKVDHTEGYNLVTIOSFKAEFR	1343
Qy	1085	IALGPKRGPSVTEKVELTCI-----KAEAA-----LCOEEQEVKAIEN	1116
Db	1344	LAGE-----VNLPKIIDCVSGDKERRQLVGRDDLQODAVMQVQPMQCNLTLLQRNTET	1397
Qy	1117	NAMVLSACVOKSALTQHRG-----KPIELSGEALDPLFMDPDPLAYTYTGSCGHVM	1168
Db	1398	RKRKLITCTYKVVPLSORSQVLEWCTGTVPJ---GEF---LVANEDGAHKRYRPNDPSAF	1451
Qy	1169	HAVCWQKYFEAVOLSSOORTHVDLFDLESEYLCPLCKSLCNTVPIPIPLQPKKINSENA	1228
Db	1452	Q---CQKMWVEQKSEKKEVEFM-----DYCQNFQPVF-----RY	1485
Qy	1229	DALAQLTLTARWITQTVLA-----RISGYNIRHAKENPIPIFFMQMGDSTLEPHSI	1280
Db	1486	FCMEKFLDPAIWEKRLAYTRSVATSIYCYL-----GLGDRHVQ-----1526	
Qy	1281	LSFGVSESIKYSNIKEMWILFATTYIRGLKVPDDRDP-----RVPMILT	1326
Db	1527	-----NILINQSAELVHIDLGVAPEOGKIPLTPETVFRLTRDIVDGMGITGVGEVF	1579
Qy	1327	WSTCAFTIOAIEN-----LLGDECKPLFGALONQHNGLKALMQFAVAQRITCP	1375
Db	1580	RRCEKTEMYWRNSQETLLITVIEVLLYD---PLF---DWTMPKLKALY---LOQR---P	1626
Qy	1376	QVLIQHLVRLLSVLPFNKISDPTCILLSIDLFLHVLVGAFLAFPSLWDXDDVDLPQSSVS	1435

```

:      : |      :      :      :      :      :      :      :      :      :
1627 EDETEL-----PTLNADQECKRNI-----SDID 1651

1436 SSYNHLYLFHLITMAHMLQILLTVDGLPLA---QVQ---EDSEEASASSFFAEISQYT 1489
:      :      :      :      :      :      :      :      :      :
1652 QSFDKAERVLMRLQELK--GVSECTVLSGCGQYNLLIQOAIIDPKNLSRLF----- 1701

1490 SGSIGCDIPGWYLMW 1504
||| ||
1702 -----PGWKAWV 1708

```

Search completed: September 25, 2003, 14:46:02
Job time : 37 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 25, 2003, 14:43:11 : Search time 54 Seconds
(without alignments)
4858.639 Million cell updates/sec

Title: US-09-724-126a-19

Perfect score: 9141

Sequence: 1 AMEGNMADEAGGTERMEIS.....ETASQETNQLFGFNQALL 1734

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 566894 seqs, 151307093 residues

Total number of hits satisfying chosen parameters: 566894

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pap.*
- 2: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pap.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pap.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pap.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pap.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pap.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pap.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pap.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pap.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pap.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pap.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pap.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pap.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pap.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pap.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pap.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pap.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	2612.5	28.6	1109	10	US-09-529-063-55
2	2612.5	28.6	1109	12	US-10-414-378-55
3	250	2.7	247	15	US-10-050-704-110
4	184	2.0	1979	15	US-10-205-823-419
5	180	2.0	996	9	US-09-815-242-5251
6	180	2.0	1009	9	US-09-815-242-12141
7	175	1.9	2383	15	US-10-082-830-260
8	168.5	1.8	3899	15	US-10-171-311-4
9	168.5	1.8	3907	15	US-10-171-311-2
10	168.5	1.8	3917	15	US-10-171-311-8
11	168.5	1.8	3925	15	US-10-171-311-6
12	165	1.8	2649	12	US-10-205-219-169
13	164.5	1.8	2139	9	US-09-727-384-6
14	164.5	1.8	2139	15	US-10-023-219-4
15	156.5	1.7	2710	15	US-10-011-366-6

16	154.5	1.7	2789	9	US-09-801-574-57
17	150	1.6	2125	10	US-09-919-172-29
18	149.5	1.6	1388	15	US-10-146-473-82
19	148	1.6	2871	15	US-10-146-473-41
20	146.5	1.6	1855	15	US-10-177-293-315
21	145.5	1.6	677	9	US-09-745-763-168
22	143.5	1.6	2835	10	US-09-885-535-4
23	142	1.6	856	10	US-09-882-529-4
24	140.5	1.5	821	10	US-09-882-529-3
25	140.5	1.5	1115	12	US-10-335-711-2
26	140	1.5	1993	14	US-10-098-979-2
27	137.5	1.5	1938	15	US-10-171-311-164
28	137.5	1.5	1945	11	US-09-927-597-2
29	137.5	1.5	1972	15	US-10-171-311-162
30	137.5	1.5	1979	11	US-09-927-597-4
31	135.5	1.5	1810	10	US-09-882-529-2
32	135	1.5	805	11	US-09-820-843A-73
33	134	1.5	1269	12	US-10-342-224-62
34	134	1.5	2568	9	US-09-866-108-3
35	133.5	1.5	3056	8	US-08-984-090-2
36	133.5	1.5	3056	11	US-09-825-476-3
37	133.5	1.5	3056	12	US-10-361-562-2
38	133.5	1.5	3056	15	US-10-307-077-1
39	133.5	1.5	3056	15	US-10-059-585-41
40	133.5	1.5	3057	15	US-10-175-225-3
41	133	1.5	1394	9	US-09-864-761-36118
42	132.5	1.4	1133	15	US-10-309-851-14
43	132.5	1.4	1133	15	US-10-309-851-16
44	132.5	1.4	1135	15	US-10-309-851-12
45	132.5	1.4	1591	15	US-10-197-666A-92

ALIGNMENTS

RESULT 1

US-09-529-063-55
; Sequence 55, Application US/09529063
; Patent No. US20020102542A1
; GENERAL INFORMATION:
; APPLICANT: FUKUSHIMA, DAIKICHI
; APPLICANT: SHIBAYAMA, SHIRO
; APPLICANT: TADA, HIDEAKI
; TITLE OF INVENTION: POLYPEPTIDE, CDNA ENCODING THE POLYPEPTIDE, AND USE OF
; TITLE OF INVENTION: THE BOTH
; FILE REFERENCE: 058769
; CURRENT APPLICATION NUMBER: US/09/529,063
; CURRENT FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: PCT/JP98/04514
; PRIOR FILING DATE: 1998-10-06
; PRIOR APPLICATION NUMBER: JP 9-274674
; PRIOR FILING DATE: 1997-10-07
; NUMBER OF SEQ ID NOS: 117
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 55
; LENGTH: 1109
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-529-063-55

Query Match	28.6%	Score	2612.5	DB	10	Length	1109
Best Local Similarity	45.9%	Pred. No.	1.4e-224	Mismatches	343	Indels	59
Matches	517	Conservative	208	Mismatches	343	Indels	59
Gaps	19						
QY	649	VLVEYPLRLVLVAQVVAEMRRNGLSLISQVFFYQDKREEMWDKDIIMLQIGASLMD	708				
Db	1	MLIEHPRLRLVLCAQVHAGMWRNRGFSLVNIIYYHNKCRREMFDKDVVLMQTGVSMMD	60				
QY	709	PNKFLILLVLORYELAEAFN-----KTISTK--DODLLKQYNTLIEEMLOVLIYIGERYV	761				
Db	61	PNHFLIMLSRFLYQIFSTPDYKGRFSSEITHDKVQOONNTLIEEMLYLILYIGKRF	120				
QY	762	PGVGNVTKEVTMRIEIIHLCLIEPMHSAITAKNLPENNETGLNVINKVATFKPGVS	821				

Db 121 PGVQVNDATDEIKREIIHQISIKPMAHSELVKSLPEDENKETGMSVIEAVAHFKKGLT 180
Qy 822 GHGVYELKDSKDFNMVYFYHYKTHQSKAHHQKRRKQKQENKDEALPPPPPPFCFPAFS 881
Db 181 GRGMVELKPECAKEFNLYFYHFSRAEQSKAAEQKRLKRNREDTALPPVLPFCPLFA 240
Qy 882 KVINLNCDDIMYILRTVFERAIDTDSNLWTEGMLQMAFHILALGLLEEKOQLKAPBEE 941
Db 241 SLVNILQSDVLMCTMGITLOWAVEHNGYAWSESMLQRLVHLHIGMALQEEKHLENVTEH 300
Qy 942 -VTFDFYHKASRLGSSAMN---IQMLEKLGIPOLEGQKDMITWILQFDTVKRLREKS 997
Db 301 VVTFFTQKISKGEAPKNSPSILAMLETQNAPLYEVHDMIRWILKTNFAVAKKRESS 360
Qy 998 -CLIVATTSGSESIKNDITHDKAEKRRKAEARLHROKIMAQMSALQKNFETHKLM 1056
Db 361 PTPSPAETEGTIM---EESRDKDAERKRAEALRLRREKIMAQMSQEMQHFIDENKEL 417
Qy 1057 YDNTSEMPGKEDSIMEEESTPAVSDYSRIALGPKRGPSVTEKEVLTICILQOEQEVKIE 1116
Db 418 FOOTLELDASTSAVLDH--SPVASDMTLTALGPAQTVPEQRFVTCILQOEQEVKVES 475
Qy 1117 NAMVLSACVOKSTALTQHRKPIELSGEALDPLFMDPDLAYGTYTSGGHVMAHVCWQY 1176
Db 476 RAMVLAFAVQKSTVLSKNRSKFIQ-DPEKYDPLFMHPDLSCGTHSSCGTHIMHAHCWQY 534
Qy 1177 FEAVOLSSQO-----RIHVDLFDLESGEVLCPLCKSLCNTVPIPILOPQKINSENADAL 1231
Db 535 FDSVQAEQRQORQLRLHTS-YDVENGFEFLCPCECLSNVPIPL--LSPRNFNNRLN-F 591
Qy 1232 AQLTLARWOTVLARISGYNIRHAKENPIPIFFNOGMGSTLFEHSILSFGVSSIKY 1291
Db 592 SDOPNLTOWIRTISQIKALQFLRKEESTPNNASTKNSENVDELQLPFGPRDPKIPY 651
Qy 1292 SNSIKEMVILFATTYRIGLVKPPDRDPVPMLTWTSTCAFTIOAENLGDGKPLFGA 1351
Db 652 SESIKEMLTTFGTATYKVLGVHPNEEDPRVIMCWGSCAYTQISIERILSDEDKPLFGP 711
Qy 1352 LQNRQHGLKALMQFAVQRTCPQVLOIQRHLVLLSVLPNIKSEDTPCLLSLDLRHVL 1411
Db 712 LPCRDLCLSLTRFAAAHWTVASVVOGHFCLFASLPVNDSSHEELPCILDIDMHLL 771
Qy 1412 VGAVLAPPSLYWDDPVDLQPSVSSSYNNHLYLFLHITMAHMLQILL---TVDTGILPAQV 1468
Db 772 VGLVLAEPALQCD----PSGISLGTGDLHIFHLVTHMAHIIQILLTSCTEGMD--QE 824
Qy 1469 OEDSEEAHSASSFAEISQVTSIGCDIP-GWYLVWSLKNIGITPYLRCAALFHYLLGV 1527
Db 825 NPPOEESAVLALYKTLHQYT-GSALKEIPSGWHLWRSVRAGIMPFLKCSALFFHYLNGV 883
Qy 1528 TPPEELHTNSAGEYSALCSVLSLPTNLFLFQEWYDVTVRPLLRRCADPALLNCKOKN 1587
Db 884 PSPDIOV-PGTSFHEFLCSVLSPLNNLCLFQENSEIMNSLIESWCNRSEVKRYLEGER 942
Qy 1588 TVVRYPKRNSLIELPDYSCLLNOASHFRCPRSADDERKHPVLCFLCGAILCSONICCO 1647
Db 943 DAIRYPRESKNLNPEDYSLINQASNFSPCKSGGKSRAPTLCLVCGSLCLCSQSYCCQ 1002
Qy 1648 EIVNGEEVGACIFHALHC-----KARGCAYPAPYLDYEGETDPLG 1687
Db 1003 TELGEDVGACHTATYSCGSGVGIFLVRRECQVFLAGKTKGCFSPYLDYGETDQGL 1062
Qy 1688 KRGNPLHLRSERYRKLHVWQOHCIIIEIARQSQTNQMLFGFNWOLL 1734
Db 1063 RRGNPLHLCKERFKKIQKLHQHVSVEIGHAQEAQNTLVGIDWQHL 1109

RESULT 2

US-10-414-378-55
; Sequence 55, Application US/10414378
; Publication No. US20030165981A1
; GENERAL INFORMATION:

; APPLICANT: FUKUSHIMA, DAIKICHI
; APPLICANT: SHIBAYAMA, SHIRO
; APPLICANT: TADA, HIDEAKI
; TITLE OF INVENTION: POLYPEPTIDE, CDNA ENCODING THE POLYPEPTIDE, AND USE OF
; TITLE OF INVENTION: THE BOTH
; FILE REFERENCE: Q58769
; CURRENT APPLICATION NUMBER: US/10/414, 378
; CURRENT FILING DATE: 2003-04-16
; PRIOR APPLICATION NUMBER: US/09/529, 063
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: PCT/JP98/04514
; PRIOR FILING DATE: 1998-10-06
; PRIOR APPLICATION NUMBER: JP 9-274674
; PRIOR FILING DATE: 1997-10-07
; NUMBER OF SEQ ID NOS: 117
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 55
; LENGTH: 1109
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-414-378-55

Query Match 28 6%; Score 2612.5; DB 12; Length 1109;

Best Local Similarity 45.9%; Pred. No. 1.4e-224; Indels 59; Gaps 19;
Matches 517; Conservative 208; Mismatches 343;

Qy 649 VLVVEPLRCLVLVAQVVAEMRRNGLSLISQVYFYQDVKREEMDYDKDIIMLQIGASLMD 708
Db 1 MLIEHPLRCLVLCAQVHAGMWRNGLSVNQIYVYHNVKCRREMPDKVVMQLTGVSMMD 60
Qy 709 PNKFLLLVQRYELAEAFN----KTISTK--DQDLIKQYNTLIEMLQVLIYIGERVY 761
Db 61 PNHFLMILSRPELYQIFSTPDYGRFSEITHKDVQVQNNLTIEEMLYLIIMLVGERES 120
Qy 762 PCVGQVNTKEEVTMRTEIHLCTIEPMPHSAIAKNLPENNNENNGLENVINKVATFKKPGVS 821
Db 121 PGVQVNDATDEIKREIIHQISIKPMAHSELVKSLPEDENKETGMSVIEAVAHFKKGLT 180
Qy 822 GHGVYELKDSKDFNMVYFYHYKTHQSKAHHQKRRKQKQENKDEALPPPPPPFCFPAFS 881
Db 181 GRGMVELKPECAKEFNLYFYHFSRAEQSKAAEQKRLKRNREDTALPPVLPFCPLFA 240
Qy 882 KVINLNCDDIMYILRTVFERAIDTDSNLWTEGMLQMAFHILALGLLEEKOQLKAPBEE 941
Db 241 SLVNILQSDVLMCTMGITLOWAVEHNGYAWSESMLQRLVHLHIGMALQEEKHLENVTEH 300
Qy 942 -VTFDFYHKASRLGSSAMN---IQMLEKLGIPOLEGQKDMITWILQFDTVKRLREKS 997
Db 301 VVTFFTQKISKGEAPKNSPSILAMLETQNAPLYEVHDMIRWILKTNFAVAKKRESS 360
Qy 998 -CLIVATTSGSESIKNDITHDKAEKRRKAEARLHROKIMAQMSALQKNFETHKLM 1056
Db 361 PTPSPAETEGTIM---EESRDKDAERKRAEALRLRREKIMAQMSQEMQHFIDENKEL 417
Qy 1057 YDNTSEMPGKEDSIMEEESTPAVSDYSRIALGPKRGPSVTEKEVLTICILQOEQEVKIE 1116
Db 418 FOOTLELDASTSAVLDH--SPVASDMTLTALGPAQTVPEQRFVTCILQOEQEVKVES 475
Qy 1117 NAMVLSACVOKSTALTQHRKPIELSGEALDPLFMDPDLAYGTYTSGGHVMAHVCWQY 1176
Db 476 RAMVLAFAVQKSTVLSKNRSKFIQ-DPEKYDPLFMHPDLSCGTHSSCGTHIMHAHCWQY 534
Qy 1177 FEAVOLSSQO-----RIHVDLFDLESGEVLCPLCKSLCNTVPIPILOPQKINSENADAL 1231
Db 535 FDSVQAEQRQORQLRLHTS-YDVENGFEFLCPCECLSNVPIPL--LSPRNFNNRLN-F 591
Qy 1232 AQLTLARWOTVLARISGYNIRHAKENPIPIFFNOGMGSTLFEHSILSFGVSSIKY 1291
Db 592 SDOPNLTOWIRTISQIKALQFLRKEESTPNNASTKNSENVDELQLPFGPRDPKIPY 651
Qy 1292 SNSIKEMVILFATTYRIGLVKPPDRDPVPMLTWTSTCAFTIOAENLGDGKPLFGA 1351
Db 652 SESIKEMLTTFGTATYKVLGVHPNEEDPRVIMCWGSCAYTQISIERILSDEDKPLFGP 711


```
Db 319 NKLSSAENDRDLRREQEQLNVEK-----RQIM-----E 348
Qy 571 ECHAVMRCS---TSFTSSKTVVQSCGHSLETYSRVSEDLVSIH--LPLSRTLGLHV 625
Db 349 ECENLKLECSKLOPSAVKQSDTM-----TEKERILQASASVEEVFRLOQALSDAEN 399
Qy 626 RLRLGAVSRHLHEFVSFEDQVEVLVEYPLRLCLVLAQVVAEMRRNGLSLSISQVFIYQ- 684
Db 400 EIMRLSSLNQDNLAE--DNLKLMRIEVLKEKSLLSQKEEL-----QMSLLKLNNVEY 454
Qy 685 -----DVKCREMYDKDILMIQIGASLMDPNKFL-----LLVLQRYEL-----AEAFNKT 729
Db 455 IKSTATRDISLDELHD---LRLMLEAKEQLNLSISEKETLIAIEIELDRQNOQATKHM 511
Qy 730 ISTKDQDLKOYN---TLBEMQLVLIYVGERVYVPGVGNVTKREVTMRIHLLCIEPM 786
Db 512 ILIKDO--LSQONEGDSIISKQDL-----NDEKKRV-----HQLEDDKM 551
Qy 787 PHSAIAKNLPENENNETGLENVINKVATFKPGVSGHGVVELKDESKLPNNFYHYHYSKT 846
Db 552 D---ITKELDVQKEKLQSEVALNDLHLTKQK-----LEDKVENLVD-----QLNKS 595
Qy 847 QHS-----KAEMOKKRRKQENKQDALPPPPPEFCPAFSKVINLLNCDIMMYILRTYFER 902
Db 596 QESNVSTQENLELKEHIRENEE-----LSRIRNEL-----MQ 629
Qy 903 AIDTDSNLWTEGLQMAFHILALCLLEKQOLOKAPE--EVTDFYHKASRLGSSAMNI 960
Db 630 SLNQDSNSNFKDTLLKREAEFVRNLKQNLSELEQNLNKKVAFDVRMENEKVLVACEDV 689
Qy 961 QMLEK--LKGIPOLEGQKMITWTLOMPD-----TVKRLREKSLIVATTSGSESI 1010
Db 690 RHOLEECLAGNQLSLEKNTIVETLKMKEGIEALCWAKKRLLEAKNKYEKTTIELSNA 749
Qy 1011 KN-----DEITH-----DKEAKRKKAE-----AARLHRQKIMQAM 1042
Db 750 RNLNTSALOHEHLKLNOKKDMELKKNIEQMDTDHKTETKDVLSLSEEQKQLTQL 809
Qy 1043 SALQKNFTIEHKLMDYNTSEMPGK-----EDSTIMEFES 1075
Db 810 INKKEIEFKERSSKLOEELDKYSQALRKNEILRTIETBKDRSLGSMKEENHLOQEL 869
Qy 1076 TPAVSDYSRIA--LGPGRGPSVTE--KEVLTCLQEBEQEVKVIENNAMVLSACVOKSTAL 1131
Db 870 ERLREEQSRAPADPKTLDSTVTELASEVSQNLTIKEHLEBEIKHHQKIIEDQNGSKMQL 929
Qy 1132 TOHRGKPIELSGEALDPLFMDPDLAYGTYTSCGHVHVAVCWQYFEAVQLSSQORIHVD 1191
Db 930 LQ-----SLQEOKKEMDEF-----RY-----QHEOMNATHQ 956
Qy 1192 LFDLESGEYLCLPLCKSLCNTVIPI---IPLQPOKINSENADALQALLTLARWIOVLARI 1248
Db 957 LF-LEKDEEI---KSQKTIQIQLTQHEERQDIQDINSDFQE-----TKV 999
Qy 1249 SGYNIRHAKGENPIPIFNQMGDSTLEFHSLSGFVSSSTKYSNIKEMVILFATTIYR 1308
Db 1000 QSLNIENG-----EKHLSKAETERLVK-----1023
Qy 1309 IGLAVPDERPRVPLMTWTSTCAFTIOAENLLGDEGKPLFGALQNR 1355
Db 1024 -GIR-----ERELEIKLLNEKNISLTQK-IDQLSKDEVGKLTQIIQOK 1064
```

RESULT 5

```
US-09-815-242-5251
; Sequence 5251, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
```

```
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5251
; LENGTH: 996
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5251
```

Query Match 2.0%; Score 180; DB 9; Length 996;

Best Local Similarity 19.4%; Pred. No. 2.8e-06;

Matches 223; Conservative 183; Mismatches 410; Indels 332; Gaps 55;

```
Qy 335 DFRQIFCQACLREEDSDENPCLIS-----RLMLWDAKLY---KCARLILHELIFSFF 384
Db 20 DFSKI-----DNNELFLISGKTGSGKTMIFDAMTYALFGKASTQREENDLRSHF 69
Qy 385 MEMEYKFLFAMFV---KYYK--OLOKEYISDDHRSISITALSVOMFTVPTL--ARHLIE 438
Db 70 ADGKQPMSTVTFEQLNHRILYKVRQGYIKEGN-----TTKNAKDFVEMVDGKYEIR 123
Qy 439 EQNVISVITETLLEVLPEYLDNRNNKFNFOGYSODKLGRIYAV--ICDLKYLILSKPIWTE 497
Db 124 ESKVIG--TQFIELL-----GVNADQFRLQFILPQGEFKRFLISNREKQG 169
Qy 498 RLRMOF---LEGFRSFLAKILTCMCGMEELRQVGQHIIEVDPDWEAAIAIOMOLKNILLM 553
Db 170 ILRTLFDSEKFEAIRILA-----EEVKKEKAO-----IENRYQOIDL 208
Qy 554 FOEWACADEE---LLLVAKECHKAVMRCSTSFSSSKTVVQSCGHSLET--KSYRVSE 607
Db 209 WQIEISFDDDKIKGLLEVATQIDKVIENIPL--LOARSKEILAFVNESKETAIKEYEIE 267
Qy 608 DLVSIHLPLSRILAG--LHVRSLRGAVSRHLHEFVSFEDQVEV--LVEYPLRLCLVLAQV 664
Db 268 -----KKTLENNILKDNINQLN--KNKIDFVLQKEQPEIEETEARLKLQDITNL 316
Qy 665 VAEWMRRNGLSL-----ISQV-FYQDVKCREEMYDKDILMIQIGA-----SLMD 708
Db 317 LNYIENREKTIETKIASKKDISETNNKILNLDCKRKNIDKKEKMLEBNGDLISKISFID 376
Qy 709 PNKFLLLVLQRYELAEAFNKITSTKQDQLIKQVNTLIEEMQLVLIIVIGRYVPGVGNVT 768
Db 377 KTRVLENDINKYQOSVYLNIRLERLFTDEQLGDELNNLIKGL-----EKVEDSIGN-- 425
Qy 769 KEEVTMRIEHLICIEPMPHSAIAKNLPENENNETG-----LENN-----LEN 807
Db 426 -NESDYEKIIEI-----NNAITNINNEINVIKENEKAKDELDKLLGSKOBLN 472
Qy 808 VINKVATFKKPGVSGHGVVELKDESKLPNNFYHYHYSKTQ---HSAEKHMQKRRKQENK 864
```

Db 473 QINEEKTI-----LKNLEIKLDYDYSKLDLNDKSFISEIKSAVKIG 515
QY 865 DEALPPPPPEFCPAFSKVINLLNCDIMVILRTVFERAIDTDSNLTGMLQAFHILA 924
Db 516 DQ-----CPICGNEIQLDGHIDFDSIAKRONKEIEIANIHTMES--NIAVHSE 564
QY 925 LGLLEEK--QOLQKAPEEVTFDYHKASRLGSSAMNIQMLLEKLGIPQLEGQKDMITWI 983
Db 565 IKFVNEKISINIKTQSDLSLEVLNKRLLLENENALNQDLNKF--IEQMKEDNLT-- 620
QY 984 LQMFDTVKRLRE-----KSC--LIV--ATTSGSESTKNDIEITHDKKAEKRKAARLH 1034
Db 621 LOIHNKQLRNKNESELKICRNLTIFETLSKYNNITNFEVDYKKYIQDVNQHOEHSNQI 680
QY 1035 ROKIMAQMSALQKNFIETHKL-----MYDNTSEMPGKEDSIMEEESTPAVSD----- 1081
Db 681 EDKLI--QLS--QRKLIQNNLHNYENQLETYNNDLEL--NEQSIEMMSRLNTDDNDIN 735
QY 1082 -----YSRIALGPKRGPSTV--EKEVLTCLCOEEQEVK----- 1113
Db 736 EIIANRGEQEELEQRDITYKKRYHEFEMEIAERLESITKDELDSKLDYELKKGKMN 795
QY 1114 --IENAMVLSAC--VOKSTALTOH-----RGRPIELSGE 1144
Db 796 TLIDEVSAVHYQCONNIKTQSTIVSHINYLQELKDQOEIQLAEITLSCGKNNKNTLENF 855
QY 1145 A-----LDPLFMPDPLAIGTYTSCGHVHMAVCWQKFEAVQ--LSSQQRIHVDLFDLESG 1198
Db 856 VLIYYLDQIAQANRLATMSDNRYQLIRR-----EAVSHGLSG--LEIDVFDLHNS 905
QY 1199 EYLCPLCKSLCNTVPIIPLQPKINS--ENADALAQLLTLARWIOVRLARISYNIHAK 1257
Db 906 -----KSRHISLSGGGETFQSSALALGSLSIQVQOQSG----- 938
QY 1258 GENPIPIFFNOGMDSTLEFHSILSGVESSIKYSIKEMVLTATTVIRIGLKVPDPE 1317
Db 939 GISLESIFIDEFG--TLDOETL-----ETVLDITLLNKSTGRWGIISHVSE 984
QY 1318 RDRPVML 1325
Db 985 LKNRIPLV 992

RESULT 6
US-09-815-242-12141
; Sequence 12141, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12141
; LENGTH: 1009
; TYPE: PRF
; ORGANISM: Staphylococcus aureus
US-09-815-242-12141

Query Match 2.0%; Score 180; DB 9; Length 1009;
Best Local Similarity 19.4%; Pred. No. 2.9e-06;
Matches 223; Conservative 183; Mismatches 410; Indels 332; Gaps 55;
QY 335 DFRIFCOACUREPDSNPCLIS-----RLMLDAKLY-----KGARKILHELIFSSFF 384
Db 20 DFSKI-----DNNLEFLISGKTGKTMTIFDAMTYALFGKASTQREENDLRSHF 69
QY 385 MEMEYKKLFAMEFV--KYYK--QLQKEYISDDHRSISITALSVQMFVPTL--ARHLIE 438
Db 70 ADGQKPSVTFEFOLNHRYYKVHRQGPYIKGN-----TTKNAKFDVFEMVDGKYEIR 123
QY 439 EONVSVITETLLEVLPEYLDNRNKNFQGYSDKLGRVAV--ICDLKYLILSKPTIWE 497
Db 124 ESKVISG--TQPIIEL-----GVNADQFROLFILPQGEFRFLISNREKOG 169
QY 498 RLRMQF----LEGFRSEFKILTCMQGMEIEIRQVGQHIIEVDPDEAAIAIQMLKNILM 553
Db 170 ILRTLDFSEKFEAREILK-----EEVKKEKAQ-----IENRYQOJDLL 208
QY 554 FQEWACDDE--LLLVAYKECHAVMRCSTSFSSSKTVVQSCGHSLET--KSYRVSE 607
Db 209 WOETIESFDDDKIKGLLEVATQIDKVIENIPL--LQARSKEILAFVNESKETAKEYEITE 267
QY 608 DLVSIHLPLSRTLAG--LHVLSRLGAVSRHLHEFVSFDFQEV--LVEYPLRCLVLVAQV 664
Db 268 -----KRTLENNILKDNINQLN--KNKIDFVLKQEQPEIEIEAKLLQDITNL 316
QY 665 VAEWRRNGLSL-----ISQV--FYQDYKCREMYDKDIIIMIQIGA-----SLMD 708
Db 317 LNYIENREKTIETKIAKSKKDISETNNKTLNLDCKRNDIDKEKKMLEENGDLIESKISFD 376
QY 709 PNKFLLLVQRYELAEAFNKTISTKDDQLIKOYNTLIBEMLQVLIYVIGERYVPGVNTV 768
Db 377 KTRVLFNDINKYQOQSYLNIERLRTDEQLDELNNLIKGL-----EKVEDSIGN-- 425
QY 769 KEVVTREIITHLLCIEPMPSAIAKNLPENNETG-----LENN-----LEN 807
Db 426 -NESDYEKIIEI-----NNAITNINNEINVIKENEKAKDELDKLLSGKQELN 472
QY 808 VINKVATFKPGVSGHGVYELKDESLEKDFNMVYFYHYSKTO--HSAEHEMOKRRKQENK 864
Db 473 QINEEKTI-----LKNLEIKLDYDYSKLDLNDKSFISEIKSAVKIG 515
QY 865 DEALPPPPPEFCPAFSKVINLLNCDIMVILRTVFERAIDTDSNLTGMLQAFHILA 924
Db 516 DQ-----CPICGNEIQLDGHIDFDSIAKRONKEIEIANIHTMES--NIAVHSE 564
QY 925 LGLLEEK--QOLQKAPEEVTFDYHKASRLGSSAMNIQMLLEKLGIPQLEGQKDMITWI 983
Db 565 IKFVNEKISINIKTQSDLSLEVLNKRLLLENENALNQDLNKF--IEQMKEDNLT-- 620
QY 984 LQMFDTVKRLRE-----KSC--LIV--ATTSGSESTKNDIEITHDKKAEKRKAARLH 1034
Db 621 LOIHNKQLRNKNESELKICRNLTIFETLSKYNNITNFEVDYKKYIQDVNQHOEHSNQI 680
QY 1035 ROKIMAQMSALQKNFIETHKL-----MYDNTSEMPGKEDSIMEEESTPAVSD----- 1081
Db 681 EDKLI--QLS--QRKLIQNNLHNYENQLETYNNDLEL--NEQSIEMMSRLNTDDNDIN 735
QY 1082 -----YSRIALGPKRGPSTV--EKEVLTCLCOEEQEVK----- 1113
Db 736 EIIANRGEQEELEQRDITYKKRYHEFEMEIAERLESITKDELDSKLDYELKKGKMN 795

[illegible]

```
Db 808 SIKSKSDVWEKEI--EILIEENEDLKQOCI-----QLNEEIEKQRTFFFAEKN 856
Qy 385 MEMEYKFLFAMEFYKYKQLOKEY-----ISDDHRSISITALSVMFTVPTLARHLIEE 439
Db 857 FEVN-----YQELQEBYACLLKVKDDLEDSKN-----KOELEY 889
Qy 440 QNVISVITETLLEVPYLDNRN---KFNQGYSDKLGRLVYAVICDKYLILISKPTIW 495
Db 890 KSKLKALNEEL-----HLQRINPTTVKMKSSVFDEK--TFVAETLEMGEVVEKDTTEL 941
Qy 496 TERLRMQFLGFRSFLKILTCMOGMEIRROVGO--HIEVDPDWEAAIAIOMQLKNILLMF 554
Db 942 MEKLEVTREK-----LELSQRLSDLSQKQKHGEISFLNEEVKSLKOEKEQVSLRC 994
Qy 555 QEW-----CACDEE-----LLLVAYKECHKVAMRCSTSFSSSKTVVQSCG 595
Db 995 RELEIINHNRAENVQSCDTQVSSLLDGVVMTSRGAEGSVKVNKSFGESKIMVED-- 1052
Qy 596 HSLKTSKRVSEED-----LVSIHLPLSRTLGLH--VRLSRGAVSRHLHEFVSFDFQVEV 649
Db 1053 -KVSFENNTVGEESKQEOILDLHLPVSVTKESSLRATOPSENDKLOKELNVLKSEQNDLRL 1111
Qy 650 LVEYPLRCLVLVAQVVAEMRRNGLSLSQVFFYQDVVKCREMYDKDIIMLQIGASLMDP 709
Db 1112 QMEAQRICLSLVYS-----THVDQVREYME-----NEKDALCSLKEELIFA 1153
Qy 710 NKFLLLVLQOR-YELAEAFNKTISTKQDQ-----LIKQYNTLIEE-----MLQVLIYVGER 759
Db 1154 QEEKIKELQKHOLELQMTKQETGDEGKPLHLLIGLKQAVSECSYFLOTLCVSLGEY 1213
Qy 760 YVPGVGNVTKEEVTMREIHLCLTEPMHPHSAIAKNLPENNETGLENVINKVATFKKPG 819
Db 1214 YTPAL-----KCEV-----NAEDKEN-----1229
Qy 820 VSGHGVYELKDESCLKDFNMVYFYHYSKTOHSAEHMOKRRKQENKDEALPPPPPEFCPA 879
Db 1230 -SGDYISENEDPELODYRYEVODFOENMHT-----LLNKVTEYKN-----1269
Qy 880 FSKVINLLNCIDIMYILRTVTFERAIDTDSNLW---TEGMLQMAFHIALGLLEEKQOLQK 936
Db 1270 -----LLVLQTRL-----SKIWQQTDCM-----KLEFEE-----N 1296
Qy 937 APEEVT-----DFYHKASRLGSSAMNIOMLLEKLGIPOLEGQKDMITWI 983
Db 1297 LPKEETEFLSIHSQMTNLEDIVNHK-----SKLSSLDQLEKTKLEEQVQSELSLSS 1350
Qy 984 LQMFDTVKRLREKS--CL-----IVATTSGSESINKDEITHDKAERKRAEAARLHQ 1036
Db 1351 QOOLKETQNYEAEIHCLOKRLQAVSESTVPPSLPDSVVITESDA-----Q 1397
Qy 1037 KIMQMSALQKNFIETHKLMVDNTEMPGKEDSIMEESTPAVSDYSRIALGPKRGPSVT 1096
Db 1398 RTMYPGSCVKNI-----DGTIEFGS--REGVKEETNI-----VKLL 1432
Qy 1097 EKEVLTICLOQEOEVKIENNAMVLSACVQKSTALTQHRGP IELSGEALDPLFMDPDIA 1156
Db 1433 EKQY-----QEOLEEEVAKVIVSMSIAFAQOTELSR-----ISG-----1466
Qy 1157 YGTYTGSGHVMHVCWKQ--YFEAVQLSSQRIHDLF---DLSGEYLCPLCKSL 1208
Db 1467 -GKENTASSKQAHAVCCQEQHYFNEMKIL--SQDQIGFTFTVDVRFKEEFKPLSKEL 1521
```

RESULT 10

```
US-10-171-311-8
; Sequence 8, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
```

```
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoersh, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 3917
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-171-311-8

Query Match 1.8%; Score 168.5; DB 15; Length 3917;
Best Local Similarity 18.7%; Pred. No. 0.00034;
Matches 201; Conservative 165; Mismatches 364; Indels 347; Gaps 50;

Qy 208 EMTINEEEKELPELOIREKNERYCVLFNDEHSHDVHVIYSQRLDCELAEOHLHTTA 267
Db 716 EMTL--QINELQOEIILROEEK-----EKGTEQEVOELQ--LKTELLEKQMK--- 760
Qy 268 IDEKERRAYKAGAYAAQCAEKEDIKSHSENVSQHPLHVEVLHSEIMAHQFALRLGSMWN 327
Db 761 -EKENDLQEKFAQLEAENSILKDEKKTLEDMLK-----IHPVSOEERLI-----FLD 807
Qy 328 KIMSYSDF---RQIFQACLRPEEPSENPCLLSRMLMDAKYLGARKILHELIFSSFF 384
Db 808 SIKSKSDVWEKEI--EILIEENEDLKQOCI-----OLNEETEKQRTFFFAEKN 856
Qy 385 MEMEYKFLFAMEFYKYKQLOKEY-----ISDDHRSISITALSVMFTVPTLARHLIEE 439
Db 857 FEVN-----YQELQEBYACLLKVKDDLEDSKN-----KOELEY 889
Qy 440 QNVISVITETLLEVPYLDNRN---KFNQGYSDKLGRLVYAVICDKYLILISKPTIW 495
Db 890 KSKLKALNEEL-----HLQRINPTTVKMKSSVFDEK--TFVAETLEMGEVVEKDTTEL 941
Qy 496 TERLRMQFLGFRSFLKILTCMOGMEIRROVGO--HIEVDPDWEAAIAIOMQLKNILLMF 554
Db 942 MEKLEVTREK-----LELSQRLSDLSQKQKHGEISFLNEEVKSLKOEKEQVSLRC 994
Qy 555 QEW-----CACDEE-----LLLVAYKECHKVAMRCSTSFSSSKTVVQSCG 595
Db 995 RELEIINHNRAENVQSCDTQVSSLLDGVVMTSRGAEGSVKVNKSFGESKIMVED-- 1052
Qy 596 HSLKTSKRVSEED-----LVSIHLPLSRTLGLH--VRLSRGAVSRHLHEFVSFDFQVEV 649
Db 1053 -KVSFENNTVGEESKQEOILDLHLPVSVTKESSLRATOPSENDKLOKELNVLKSEQNDLRL 1111
Qy 650 LVEYPLRCLVLVAQVVAEMRRNGLSLSQVFFYQDVVKCREMYDKDIIMLQIGASLMDP 709
Db 1112 QMEAQRICLSLVYS-----THVDQVREYME-----NEKDALCSLKEELIFA 1153
Qy 710 NKFLLLVLQOR-YELAEAFNKTISTKQDQ-----LIKQYNTLIEE-----MLQVLIYVGER 759
Db 1154 QEEKIKELQKHOLELQMTKQETGDEGKPLHLLIGLKQAVSECSYFLOTLCVSLGEY 1213
Qy 760 YVPGVGNVTKEEVTMREIHLCLTEPMHPHSAIAKNLPENNETGLENVINKVATFKKPG 819
Db 1214 YTPAL-----KCEV-----NAEDKEN-----1229
Qy 820 VSGHGVYELKDESCLKDFNMVYFYHYSKTOHSAEHMOKRRKQENKDEALPPPPPEFCPA 879
```

Db 1230 -SGDYISENEDPELDQYRYEVQDFQENMHT-----LLNKYTEYNN----- 1269

Qy 880 FSKVINLLNCDIMYILRTVFERAIDTDSNLW---TEGMLQMAFHILALGLLEEKQOOLQ 936

Db 1270 -----LLVLQTRL-----SKIWQOOTDGM-----KLEFEGEE-----N 1296

Qy 937 APEEEVTF-----DFYHKASRLGSSAMNIQMLLEKLGIPQLEGOKDMITWI 983

Db 1297 LPKETEFLSHSQMTNLEDIDVNHK-----SKLSSQDLKTKLEEQVQLESLSL 1350

Qy 984 LQMFDTVKRLREKS--CL-----IVATTSGSESINKDEITHDKKAERKKAERKKAERLHRQ 1036

Db 1351 QOOLKETEQNYEAEIHCLOKRLQAVSESTVPPSLPVDVSVITESDA-----Q 1397

Qy 1037 KIMAQMSALQKNFIETHKLMYDNTSEMPGKEDSIMEESTPAVSDYSRIALGPKRGPSVT 1096

Db 1398 RTMYPGSCVKKNI-----DGTIEFSG-EFGVKEETNI-----VKLL 1432

Qy 1097 EKEVLTCILCOEQEVEKIENNAVMVLSACVOKSTALTOHRGKPIELSGEALDPLFMDPDLA 1156

Db 1433 EKQY-----QOOLEEVAKVIVMSIAFAQOQTELSR-----ISG----- 1466

Qy 1157 YGTYTSCGHVMAHVCWK--YFEAVQLSSQORIHVDLF---DLESGEVLCPLCKSL 1208

Db 1467 -GKENTASSKQAHAVCQEQHGFENMKL-SQDQIGFQTFETVDVKFEKPELPSKEL 1521

RESULT 11

US-10-171-311-6

; Sequence 6, Application US/10171311

; Publication No. US20030087270A1

; GENERAL INFORMATION:

; APPLICANT: Schlegel, Robert

; APPLICANT: Chen, Yan

; APPLICANT: Zhao, Xumei

; APPLICANT: Monahan, John

; APPLICANT: Kamatkar, Shubhangi

; APPLICANT: Glatt, Karen

; APPLICANT: Gannavarapu, Manjula

; APPLICANT: Hoerish, Sebastian

; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF CERVICAL CANCER

; FILE REFERENCE: MRI-035

; CURRENT APPLICATION NUMBER: US/10/171,311

; CURRENT FILING DATE: 2002-06-12

; PRIOR APPLICATION NUMBER: US 60/298,159

; PRIOR FILING DATE: 2001-06-13

; PRIOR APPLICATION NUMBER: US 60/298,155

; PRIOR FILING DATE: 2001-06-13

; PRIOR APPLICATION NUMBER: US 60/335,936

; PRIOR FILING DATE: 2001-11-14

; NUMBER OF SEQ ID NOS: 238

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 6

; LENGTH: 3925

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-171-311-6

Query Match 1.8%; Score 168.5; DB 15; Length 3925;

Best Local Similarity 18.7%; Pred. No. 0.00034;

Matches 201; Conservative 165; Mismatches 364; Indels 347; Gaps 50;

Qy 208 EMTTWEKEKLPPELQIREKNERYCVLFNDEHHSHYVLSLQRLDCELAEOALHTTA 267

Db 716 EMTL-QINLEKQETELRQEK-----EKGLEGEVQBLQ-LKTELLEKQMK--- 760

Qy 268 IDKEGRRAVKAGAAQCAQEAEDIKSHSENYSOHPLHVEVLHSEITMAHQKFAIRLGSMWN 327

Db 761 -EKENDLQERKFAQLEAENSILKDEKKTLTLEDMLK-----IHTPVQSEERLI-----FLD 807

RESULT 12

US-10-205-219-169

; Sequence 169, Application US/10205219

; Publication No. US20030138803A1

; GENERAL INFORMATION:

; APPLICANT: Warner-Lambert Company

; APPLICANT: Lee, Kevin

; APPLICANT: Dixon, Alistair

```
; APPLICANT: Brooksbank, Robert
; APPLICANT: Pinnock, Robert
; TITLE OF INVENTION: Identification and Use of Molecules Implicated in Pain
; FILE REFERENCE: WL-A-018200
; CURRENT APPLICATION NUMBER: US/10/205,219
; CURRENT FILING DATE: 2002-07-24
; PRIOR APPLICATION NUMBER: GB 0118354.0
; PRIOR FILING DATE: 2001-07-27
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 169
; LENGTH: 2649
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Pemphigoid antigen
; US-10-205-219-169

Query Match 1.88; Score 165; DB 12; Length 2649;
Best Local Similarity 17.88; Pred. No. 0.00035;
Matches 349; Conservative 262; Mismatches 644; Indels 710; Gaps 92;

QY 18 EISAELPQPPORLASWMDQOVDFYTAFLHLAOLVPEIYFAEMDPD---LEKQEEVSOMS 74
DB 417 ELMRELDQKEENIKS--VQEIARQLLENHAPRLTTEAYRAAMQTOWSWILQLCQCVEQH 474

QY 75 IFTPLEWYLFGEPPDICLEKLKH-SGAFOQLCGRVFKSGETYSYSC-RDCAIDPTCVLCMDC 132
DB 475 IKENTAYFEFFNDAKEATDYLRNLKDAIQ-----RKYSRDSRSSHK-----LEDL 520

QY 133 FQDSVHKHRYKMHSTGGGFCDCGDTAEAWKTGFCVNHPEPGRAGTI-----KENSRCPLN 188
DB 521 VQESMEKEKELQYKST-----IANLMGAKATTIQLKPRNSDCPLK 561

QY 189 EEVIVQARKTFPSVIKY-VVEMTIWEEKEKELPELQIREKNERYCYVLFNDEHSHYDHI 247
DB 562 TSIPIKA-----ICDYRQIEITYKDD-----CVLANNSHRAKWKVI 599

QY 248 YSLQALDCBLAEALQHTT--AIDKEGRRRAVAGAAACOEAKEDIKSHSENVSQ--HPL 303
DB 600 SPTGN-----EAMVPSVCFTVPPNKEAV-----DLANRIEQOYQNVLTWHES 643

QY 304 HVEV-----LHSEI-----MAHQKFAIRLGSGWNKIMSYSDPRQI 339
DB 644 HINNKSVSWHYLINEIDRIASNVASIKTMLPGEHQVLSNLQSRPEDELSQE-SQV 702

QY 340 FCOACLEEDPSENCLISRLMLWDAKLYKGARKILHELIFSSFFMEME-----388
DB 703 FSGSDITQLEKEVNV-----KQYQELLKSAEREQEEVSYNYIYSEVRNIRLNCED 758

QY 389 -----YKKLFAM-EFVKYKQLOKEYISDDHRSISITALSVMQF-----427
DB 759 RLRIQRTPLERDLHESVFRITEQELKKELER--LKDDLQ---TITNKEEFFSQAAA 813

QY 428 --TVPTLARHLIEQNVISVITETLLEVLPEYLDNRNKNF--OQYSQDKLGRVY----478
DB 814 SSSVPTLR-----SELNVVLQNMNQVSMSSYTDKLTNTVLNLTQAAEALVLYETKL 869

QY 479 ----AVICDLKYI--LISKPTIWTERL-----499
DB 870 CEEAVIADKNNTIENLISTLKQMRSEYDEKRVFHALELQKAKAISDEMFKYKERDL 929

QY 500 -----RMOFLEGRPSFLK-----ILTCMOGMEETRRQ 526
DB 930 DFDHHEKADQLVERQNVHVDINRLRLEGIGKSLKYRYDTHPLDDWIQQVETQRK 989

QY 527 VGOHIEVDPPWEAAIAIQMKNLIMLFQEWCADEELLVAYKECHKAVMRCSTSFSS 586
DB 990 IQEN--Q-SPENSKTLATQLAQKMLV-----SEIEMKQSKMDECQKAEQYSA-----1034

QY 587 SKTVQSCGSHLETKSYRV-----SEDLVSHLPLSRTLGLHVLRLSR 629
DB 1035 --TVKD---YELQTMITYAMVDVSDQKSPVKRRRRMOSSADLI-----IQEFMDLRTRYTA 1083
```

```
QY 630 LGAVSRLLHEFVSPEDFQVEVLVVEYPL-RLCLVLVAQVVAEMRRNGLSLISQVFFYQDYK 688
DB 1084 L--VTLMQTVIKFAGDSLKRLEEEIKRC-----KTSSEHGAYSJDLQ 1124
QY 689 REEMYDKDIIMQLQIGASIMDPNKF-----LLILVQRYELAAEFNKTISTKDDOL 737
DB 1125 RQK-----ATVLENSKLTGKISELERMVAELKKOKSRVEELPKVRAENEL 1172
QY 738 IKQYNTLIEMLQ-VLIYIVGERYVPGVGNVTK-EVTMRI--IHLICIEPMPHSAIAK 793
DB 1173 RKQQRNVEDISLQIRAESEAKQYRRELETIVREKEAARELERVQLTIE-----AEAK 1227
QY 794 NLPENENNETGLENVINKVA--TFKKPGVSGHGVYELKDESLKDFENMYFYHYKTHQSKA 851
DB 1228 RAAVEEN---LLNFRNQLEENTFTRTLEDH--LKRKDLSLNDL-----EOKN 1271
QY 852 EHMOKKRRKROENKDEALPPPPPPFCFPAFSKVINLLNCIDIMMYILRTVFERAIDTDSNLW 911
DB 1272 KLMEELRRKRNEEELL-----KLIQOMEKDL-----AFQKV-----1304
QY 912 TEGMLQMAFHILALGLLEEKQOOLQKAPEEVT-FDFYHKASRLGSSAMNIQMLEKLKGI 970
DB 1305 -----AEQOLKEKQKIELEARRKITEIQYTCRENALPVCPTQATSCRAVTGL 1352
QY 971 POLEGQKDMITWILOMFDTVRLREKSLIYATTSSESINKD--EITHDKB--KAERKR 1026
DB 1353 QQ-BHDKQKABELKQOVDEL-----TAANKRABEQDMRELYELYNALOLEKTS 1398
QY 1027 KAAEARLHROKI-----MAQMSALQKNFIETHKLWYDNTSEMPCKEDSIMEE 1073
DB 1399 SEEKARLLKDKDLTNTNTRCLKLELRKQAEKGYSOQLRELGRQLNQTTGKAEEAQE 1458
QY 1074 ESTPAVSYSRIALGPKRSPVTEKVELTCLCOBEOEVKTIENNANVLSACVQKSTALTQ 1133
DB 1459 -----ASDLKI---KRYQL-ELESN---HEKGKLOREVDITRAHAAVEKN--IQ 1502
QY 1134 HRGPIELSGEALDPLFMDPDLAYGYTGTGSGHVMHVCWKYFPAVOLSQORITHVDLF 1193
DB 1503 HLNSQI-----HSPRDEKELERLQTCORKSDHLREQ 1533
QY 1194 DLESGEYLCPLCKSLCNTVPIPILOPOKINSENADALAQLTLARWTQTVLARISGYNI 1253
DB 1534 FEKSHQQL-----LONIKAEKENNDKIORL-----1558
QY 1254 RHAKGENPIPIFFNOGMDSTLEPHSILSFGVSESIKYNSIK-----EMVILFA 1303
DB 1559 -----NEELEKSN-ECAEMLKQKVEELTRQNETKLMQMORIAESENIVLEK 1604
QY 1304 TTIYRIGLKVPPDERPVPMLTWSTC-AFTIQATENLLGDEGKPLFGALQNRQINGLKA 1362
DB 1605 QTIQ-----RCEALKIQA-----DGFQDQLRSTNEHLHKQTKT 1638
QY 1363 LMQF-----AVAQRIT-----CPQ--VLIQ--KHLVRLSVVLPNKSIEDTPC 1401
DB 1639 BQDFQRKTKCEEDLAKSONLVSEPKQKCDQOQNIITQNTKKEVRNLNLAELNASKEE----1694
QY 1402 LLSIDLPHVLVGAFLAPSLYWDVDPDLPQSPSVSSSYNHYL-----FHLIT-----MAHML 1453
DB 1695 -----KRGCEQVQLOQAQVQELNNPLKKVQDELHLKLTIEEQWTHRK 1736
QY 1454 QILLTVDTGLPLAQVQDESEEAHSASSFFAEISQYTSIGCDIPGWYL-WYSLANGITTP 1512
DB 1737 MVLQEEESG---KFKQSAEEFRKKWEKLMESKVITEN-----DISGIRLDFVSLQ-----1783
QY 1513 YLRCAALEFFHYLLGVTPPEELHTNSAEGEYSALCSYLSLPTNLFLL---FOEYWDTVRPL 1569
DB 1784 -----QENSRAQENAKLCE-----TNIKELEROQY-----REQ 1813
QY 1570 LQRRCADPA--LLNCLKOKNTVVRPRKRNLSIELPDDYSCLLNQASHRCPRSDDERK 1627
DB 1814 MQQGHMEANHYQKQCKLEDELIAQKREVENLKKQKDDQI-----KEHE 1857
```


QY 1628 HPVLCLFCGAILCSQNICQEIIVNGEEVGACIF-----HA-----LHCKARGCAYP 1673
Db 1858 HOLVLQ-----EIOKSTAKDCTFKPDMFTVYKCOHSGELSSRNTGHLHP 1905
QY 1674 APYLDEYGETDPLKRGPNLHLSRERYRKLHLVWQOHCIIETAR 1718
Db 1906 TP-----RSPLLRWTPQPLEEKW-QHRVVEQIPK 1935

RESULT 13

US-09-727-384-6
; Sequence 6, Application US/09727384
; Patent No. US20020098511A1
; GENERAL INFORMATION:
; APPLICANT: Myriad Genetics, Inc.
; APPLICANT: Heichman, Karen
; APPLICANT: Cimbara, Daniel M.
; APPLICANT: Bush, Angie
; APPLICANT: Mauck, Kimberly
; APPLICANT: Bartel, Paul L.
; TITLE OF INVENTION: Protein-Protein Interactions
; FILE REFERENCE: 2318-271
; CURRENT APPLICATION NUMBER: US/09727,384
; CURRENT FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: US 60/168,377
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: US 60/168,379
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: US 60/185,056
; PRIOR FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 2139
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-727-384-6

Query Match 1.8%; Score 164.5; DB 9; Length 2139;
Best Local Similarity 16.7%; Pred. No. 0.00027;
Matches 241; Conservative 227; Mismatches 510; Indels 461; Gaps 54;

QY 174 GRAGTIKE--NSRCLNEEVIVQARKIFPSVIVKIVVEMTWEEKELPPELOIREKNR 230
Db 687 GOAAVLKKAHHEATCRHEE-----KKQLQVKLEEEKTHLOEKRLQHEM- 732
QY 231 YVCVLFNDEHSHYDVIYSLQALDCELAEAOHTTAIDKEGRRAVKAGAY-----AAC 284
Db 733 -----LAKRLTQAO---ASPERE-REGLOSSAWTEEKVRGLT 765
QY 285 QEAKEDIKSHEN-----VSQHPLHVEVLHSEITMAHQKFAIRLGSWNMKIMSYSDFRQIF 340
Db 766 QELEG---FHQBOLTSVLEKHTLEELRKELEKHQRELQEG----- 805
QY 341 COACLRPDSNPCLISRLMWDAKLYKGARKILHELIFSFSEFFMEMEYKILFAMEFVKY 400
Db 806 -----REKMETECNRNRSQI---EAFQSDCKQVTERCESALQSLGGRYR-----QE 849
QY 401 YKOLQKEYISDDHRSISITALSVQMTVPTLARHLIEEQNVISVITETLLEVLPE-YLD 459
Db 850 LKDOEQOREKSWQEFKEDELTOCAEAQELKLTLEKREKTSVLVLTQEREMLEKTYKE 909
QY 460 RNNKF--NFQYSQDKLGRVYAVICDLKYLISKPTIWTPE---LRMQPLEGFRGFLKIL 514
Db 910 HNSMVVERQQLQD-----LEDLRNVSTQOSLSDQILELKSSHKRLREEREVL 961
QY 515 TCMQGMEEIRRVQGOHIEVDPDWEAAIAIQMOLKNILLMFQEWACDEELL--LVAYKEC 572
Db 962 COAGASELASQRLERLEMEHDQER-----QEMMSKLLAMENI 999
QY 573 HKAVMRCSTSISSSKTVVQSGGHSLETKSVRSVSDLVSIHPLSRTLAGLHVRLSRL-G 631
Db 1000 HKAT--CETA-----DRERAEMSTEISRLQS 1023

RESULT 14

US-10-023-219-4
; Sequence 4, Application US/10023219
; Publication No. US20030032592A1
; GENERAL INFORMATION:
; APPLICANT: Myriad Genetics, Inc.
; APPLICANT: Cimbara, Daniel M.

QY 632 AVSRHLFEVDFEQVEVLVEYPLRCLVLVAQVVAEMWRRNGLSLSIQVFFYQDVKCRBE 691
Db 1024 KIKEMOQATS-----PLSMQSGQCVIGE-----BEV 1050
QY 692 MYDKDIIMLQIGASLMDPNKFLLVLTQV-YELAEAFNKTISTKODDLIKOYNLTLEMLQ 750
Db 1051 EGDGALSLLQOGBOLLEENGVLVLSQRAHQAVKENVKMATETISRLOQRLK--EPGL 1108
QY 751 VLIYIVGERYPVGVGNVTKEEVTMREITHLICIEPMPHSAIAKLNPENENNE---TGLEN 807
Db 1109 VMSSCLDEPATEPEFGNTAEQTEQFLQNRRTKQVGVTRRRHVLSLDEDEVDRLSGTGSS 1168
QY 808 VINKVATFKKPGVSGHGVYELKDESLKDFNMFYHYSKTQISKAHMKQKR-----858
Db 1169 VQRQEVKIESEASVEGSELEN-----SEETRTESWELKNQISQLOEQ 1213
QY 859 -----RKQENKDEAL-----PPPPPEPCPAFSKVINLLNC-DIMMY 894
Db 1214 MMLCADCDSASEKKQDILLFDVSVLKKLKLMLERPEASPKYKLLYEDVSRENDCLOEELR 1273
QY 895 ILRTVFERAIDTDSNLWTEGMLQMAFHILALGLLEEKQLOKAPEEEEVTFDFYHKASRLG 954
Db 1274 MMETRYDEALENNKELTAE-----VFLQDELKKMEEVTFETFLSLEKSYDEVKIE 1323
QY 955 SSAMNTQML-----LEKLGIPOLEGQKDMITW-----ILQMPDT-----989
Db 1324 NEGLNVLVLRLOGKIEKIQ--ESVVQRCDCLWEASLENLEIEPDGNILQNLQTLCECVP 1381
QY 990 -----VKRLREKSLIVATTSGSESINKDEI-----TH-----1017
Db 1382 RVRSVHHVIEECKQENQYLEGNTQLLEKVKHAHEIAWLHGTIQTQHRPRVQNVILENT 1441
QY 1018 -----DK-----EKAERKRAEAARLHRQK-----IMAO 1042
Db 1442 TLGFDQKHFOHOATTAELBETKTLQELTKLKERVTILVKQDVLVSHGKEKEELKAMM 1501
QY 1043 SALQKNFIETHK----LMYDNTSEMPGKEDSIMEEESTPAYSD--YSRIALGKRGPSVT 1096
Db 1502 HDLQITCSEMQQKVELLRYE--SEKLOQENSILRNEITTLNEEDSISNLKGLTNG---S 1556
QY 1097 EKEVLTCILCOESEQEVIENNAVLSACVOKSTALTQHRGKPIELSGALDPLFMDPDLA 1156
Db 1557 QEE-----MMOKTETVQKEN-----AAVQKMNENLKKQISELKIKNQOLD-----1596
QY 1157 YGTYTCSGHVMAVHCQKYEAVOLSSQORIHVDLFDLESGEVLC-----1202
Db 1597 -----LENTELQKNSQNOEKLOELNORLTMLCOKEKEPEGNSALEER 1639
QY 1203 -----PLCKSLCNTVIPI-----IPLQPKQKINSEN---ADALAQLTLTARW- 1240
Db 1640 EQEKNLKELEKCKVQSSTLVSSLEAESEVSKIOTHIIVQOENHLLKLEKMKQLHRCP 1699
QY 1241 -----IQTVLARIAGYNIRHAKGENPIPIFN---QGMGDSITLFSHLSIFGVE--SSIKY 1291
Db 1700 DLSDFQOKTSVLSVYNEKLLKEALSEELNSCVDKLAKSLLEHRIATMKQEQSWBHQ 1759
QY 1292 SNSIKEMVIL-----FATTIVRIGLVPPDRPRVPLMTWTSCAFTIOA---IENL 1340
Db 1760 SASLSQVASQEKVONLEDTVQNVNLOMSRMSDLRVTQOEKALQEVMSLHKQLQNA 1819
QY 1341 LGDEGRPLFGALONRQHNGLKALMQFAVORITCPQVLIQKHLVRLLSVLPNIKSEDT 1399
Db 1820 GGSWAPEIATHPISGLHNQOKRLSWDKLDHLMNEEQQLWQENERLQTMV-QNTKAE 1877

```

; APPLICANT: Heichman, Karen
; APPLICANT: Bartel, Paul L.
; TITLE OF INVENTION: Protein-Protein Interactions
; FILE REFERENCE: 2318-282-II
; CURRENT APPLICATION NUMBER: US/10/023,219
; CURRENT FILING DATE: 2001-12-20
; PRIOR APPLICATION NUMBER: US 60/256,983
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 2139
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-219-4

Query Match      1.88; Score 164.5; DB 15; Length 2139;
Best Local Similarity 16.7%; Pred. No. 0.00027;
Matches 241; Conservative 227; Mismatches 510; Indels 461; Gaps 54;

Qy 174 GRAGTICE--NSRCPLNEEVIVQAKIPFVSVKVVVEMTWEEKEELPELOIREKNER 230
Db 687 GQAALVKAHEATCRHEE-----KKLQVKEBEKTHLQEKRLQHEME- 732
Qy 231 YYCVLFNDEHSHYDHVYISLQRALDCELABQAHTTAIDKEGRRVAKAGAY-----AAC 284
Db 733 -----LKARLTQAAQ---ASFERE-REGLOSSAWTEEKVYRGLT 765
Qy 285 QEAKEDTKSHEN----VSQHPLHVEVLHSEIWAHQKFAFLRGLGSWNKIMSYSDDFQIF 340
Db 766 QELEQ----FHQELTSLVKEHTKELEKRELLEKHQRELOEQ----- 805
Qy 341 CQACLREEPSNPCLLSRLMLMDAKLYGKARKILHLELIFSSFMEMEXKKLFAMEFVKY 400
Db 806 -----REKMETECNRTSQI---EAQFSQCQKVTCECSALQSLEGYR-----QE 849
Qy 401 YKQLQKEYISDDHRSITALSQVMPVTPTLARHLTEEQNVISVITETLLEVLPE-YLD 459
Db 850 LKDIQEQOREEKQWFEKDELTOCEAEQAELLKETLKRKTTSLVLTQEREMLEKTYKE 909
Qy 460 RNNKF--NFGQYSODKLGRVAVICDLKYILISKPTTWTER---LRNQFLEGFRSFUKIL 514
Db 910 HLNSMWYVERQQLQD-----LEDLRNVSETQQSLLSDQILLEKSSHKRELREEREVL 961
Qy 515 TCMQGMEEIRQVQGHIEVDPDNEAATAIQMLKNILLMFQEWACDEELL--LVAKKEC 572
Db 962 CQAGASQOLASQRLEMEHDOQR-----OEWMSKLLAMENI 999
Qy 573 HKAVMRCSTPISSSKTVVQSCGHSLETKSVRVSEDLVSIHLPLSRRTLAGHLVRLSRL-G 631
Db 1000 HKAT--CETA-----DREAREMSTEISRLQS 1023
Qy 632 AVSKLHEFVSFEDQVEVLVEYPLRCLVLVAQVVAEMRRNGLSLSQVFFYQDVCKREE 691
Db 1024 KIKEMQQAATS-----PLSMLQSGCQVIGE-----EEV 1050
Qy 692 MYDKDITMLQIGASIMDPNFKLLLVLQV--YELAEAFNKTTSTKDQDLIKQYNTLIEMLQ 750
Db 1051 EGDGALSLLQOGEOLLEENGDVLLSLQRAHQEAVKENVKMAETISRLQQRQLKL--EPGL 1108
Qy 751 VLIIVIGERYVPGVGNVTKEVTMRETIHLLCIEPMPHSAIAKNLPENENNE---TCLEN 807
Db 1109 VMSSCLDEPATEFFGNTAEQTEQPLQONRPTQVEGVTRRHVLSDEDDVEDVLDGSTGTS 1168
Qy 808 VINKVATFKPGVSGHGYELKDESLKDFNMNFYHYSKQTHQSKAEHMQKKR----- 858
Db 1169 VQREVKTEESEAASVEGFSLEEN-----SEETRTESEWELKNQISQLQEQ 1213
Qy 859 -----RKQENKDEAL-----PPPPPEFCFAFSKVINLLNC-DIMMY 894
Db 1214 MMLCADCDSRASEKQDILLFDVSVLKKLKMLERIPEASPKYLLIYEDVRENDCLQBELR 1273
Qy 895 ILRVTFFERAIDTDSNLNTEGMQMAFHIALGLLEEKQQQKAPKEEVTFTDVFHKASRLG 954

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 25, 2003, 14:51:16 ; Search time 59 Seconds
(without alignments)
2826.381 Million cell updates/sec

Title: US-09-724-126A-19
Perfect score: 9141
Sequence: 1 AMEGNMADEEAGGTERMEIS.....EIARSQETNQMLFGFNWQLL 1734

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR.76:*

1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8480	92.8	1757	2 T14318	ubiquitin-protein
2	1865.5	20.6	1927	2 T25604	hypothetical prote
3	990	10.8	1958	2 T39808	hypothetical prote
4	866	9.5	2052	2 T37711	probable n-end-rec
5	691.5	7.6	1950	2 S12332	ubiquitin-protein
6	628.5	6.9	1941	2 T30554	ubiquitin-protein
7	606.5	6.6	1225	2 T48251	ubiquitin-protein
8	535.5	5.9	1872	2 S64851	probable membrane
9	507.5	5.6	2168	2 D88131	protein F10G7.10 [
10	292	3.2	795	2 T48252	eceriferum3 (CER3)
11	218	2.4	271	2 T40238	hypothetical prote
12	205.5	2.2	3187	2 JC5837	364K Golgi complex
13	190.5	2.1	3259	1 A56539	giantin - human
14	186	2.0	1410	1 A57013	early endosome ant
15	182	2.0	2442	2 T08621	centrosome associa
16	180.5	2.0	1413	2 T26467	hypothetical prote
17	179.5	2.0	2346	2 T13829	tptr homolog - frui
18	176	1.9	3225	2 I52300	giantin - human
19	175.5	1.9	1427	2 S22695	restin - human
20	175.5	1.9	1526	2 T41522	myosin ii - fissio
21	175.5	1.9	1830	1 S19188	myosin-v - chicken
22	175.5	1.9	2663	1 S28261	centromere protein
23	174.5	1.9	1009	2 C89910	hypothetical prote
24	174.5	1.9	1392	2 A43336	microtubule-vesicl
25	172.5	1.9	1780	2 T17272	hypothetical prote
26	168	1.8	1837	2 T41023	probable nuclear p
27	166	1.8	1163	2 G97236	ATPase involved in
28	166	1.8	1676	2 E71410	probable centromer
29	165	1.8	2649	2 A40937	bullous pemphigoid

30	164.5	1.8	1181	2 T01799	hypothetical prote
31	164	1.8	2954	2 T14156	kinesin-related pr
32	163.5	1.8	1339	2 H81307	restriction modifi
33	162	1.8	1269	2 F84730	probable myosin he
34	162	1.8	2335	2 T40186	probable phosphati
35	157.5	1.7	1642	2 T08880	NMDA receptor-bind
36	156.5	1.7	1496	2 T05634	hypothetical prote
37	156.5	1.7	1790	2 S67593	transport protein
38	156.5	1.7	2710	2 A37052	toxin A - Clostrid
39	156.5	1.7	2829	2 A42771	reticulocyte-bind
40	156	1.7	1853	1 A46761	myosin heavy chain
41	155.5	1.7	1216	2 A28821	1-phosphatidylinos
42	155	1.7	1025	2 E86355	hypothetical prote
43	155	1.7	1173	1 A53430	1-phosphatidylinos
44	154.5	1.7	2269	2 T28677	rhoptyr protein -
45	154	1.7	2469	2 H36812	hypothetical prote

ALIGNMENTS

RESULT 1

T14318
ubiquitin-protein ligase E3-alpha - mouse
N:Alternate names: N-recognin E3-alpha
C:Species: Mus musculus (house mouse)
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
R:Kwon, Y.T.; Reiss, Y.; Fried, V.A.; Hershko, A.; Yoon, J.K.; Gonda, D.K.; Sangan, P.
Proc. Natl. Acad. Sci. U.S.A. 95, 7898-7903, 1998
A:Title: The mouse and human genes encoding the recognition component of the N-end ru
A:Reference number: Z17977; MUID:98318583; PMID:9653112
A:Accession: T14318
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1757 <KWO>
A:Cross-references: EMBL:AF061555; NID:g3170886; PID:g3170887; PIDN:AAC40165.1
C:Genetics:
A:Gene: Ubql
A:Map position: 2

Query Match	92.8%	Score 8480;	DB 2;	Length 1757;
Best Local Similarity	91.0%;	Pred. No. 0;		
Matches 1598;	Conservative 64;	Mismatches 67;	Indels 28;	Gaps 4;
Qy	6	MADEEAGGTERMEISAEIPOTPORLASWDDQOVDFYTAFLHLHLAQLVPEIYFAEMDPDLE	65	
Db	1	MADEEMDGAERMDVSEPPPLAPORPASWDDQOVDFYTAFLHLHLAQLVPEIYFAEMDPDLE	60	
Qy	66	KQESVQMSITFTPLEWYLFGEDEPDICLEKLKHSQAFQICGRVFKSGGTTYSCRDCAIDPT	125	
Db	61	KQESVQMSITFTPLEWYLFGEDEPDICLEKLKHSQAFQICGRVFKSGGTTYSCRDCAIDPT	120	
Qy	126	CVLCMDFQDSVHKHNRHYKMHSTGTGGFCDCGDTGKMTGPFVFNHPEGRAGTTIKENSRC	185	
Db	121	CVLCMDFQDSVHKHNRHYKMHSTGTGGFCDCGDTGKMTGPFVFNHPEGRAGTTIKESLHC	180	
Qy	186	PLNEEVIVQARKITFPSVIKYVEMTITWEEKEKLPELQIREKNERYICYVLFNDEHSHYDH	245	
Db	181	PLNEEVIVQARRITFPSVIKYVEMTITWEEKEKLPELQIREKNERYICYVLFNDEHSHYDH	240	
Qy	246	VIYSLOALDCELAELQHTTTADKEGRRAVKAGAYACQAEKEDIKSHSENVSQHPHUV	305	
Db	241	VIYSLOALDCELAELQHTTTADKEGRRAVKAGVYATQAEKEDIKSHSENVSQHPHUV	300	
Qy	306	EVLSHSEIMAHQKFAIRLGSWNMKIMSYSSDFRIFQCOACLRPEPDSENPCILSRLLMWDA	365	
Db	301	EVLSHVVMHAHQKFAIRLGSWNMKIMSYSSDFRIFQCOACLRPEPGSENPCILSRLLMWDA	360	
Qy	366	KLYKGARKILHELIFSSFTFMEYKKFLFAMEFVKYKQLOKEYISDDHDSRISITALSVO	425	
Db	361	KLYKGARKILHELIFSSFTFMEYKKFLFAMEFVKYKQLOKEYISDDHDSRISITALSVO	420	

QY 426 MFTVPTLARHLIEQNVISVITETLLEVLPEYLDNRNKNFQGYSDQKLGRRVAVICDLK 485
Db 421 MUTVPTLARHLIEQNVISVITETLLEVLPEYLDNRNKNFQGYSDQKLGRRVAVICDLK 480
QY 486 YILISKPTIWTBLRQWFLQEGFRSFLKILTCMGMEEIRQVGOHIEVDPDWEAAIAIQM 545
Db 481 YILISKPTIWTBLRQWFLQEGFRSFLKILTCMGMEEIRQVGOHIEVDPDWEAAIAIQM 540
QY 546 QLNKILLMFQEWACDELLLVAYKECHKAVMRCSTSFISSTVTVQSGHSLTKSYRV 605
Db 541 QLNKILLMFQEWACDELLLVAYKECHKAVMRCSTSFISSTVTVQSGHSLTKSYRV 600
QY 606 SEDLVSIHLPLSTLAGLHVRLSRLGAVSRHLHFVDFEQVEVLPEYPLRCLVLVAQVV 665
Db 601 SEDLVSIHLPLSTLAGLHVRLSRLGAVSRHLHFVDFEQVEVLPEYPLRCLVLVAQVV 660
QY 666 AEMWRNGLSLISQVYQDVCKREEMDKDIIIMLQIGASIMDPNPKFLLVLQRYELAEA 725
Db 661 AEMWRNGLSLISQVYQDVCKREEMDKDIIIMLQIGASIMDPNPKFLLVLQRYELTDA 720
QY 726 FNKTISTKQDQLIKQYNTLIEEMQLVLIYVGERYPVGVGNVTKKEEYTMREIHLGLCIEP 785
Db 721 FNKTISTKQDQLIKQYNTLIEEMQLVLIYVGERYPVGVGNVTKKEEYTMREIHLGLCIEP 780
QY 786 MPHSATAKNLPENNETGLENVINKVATPKPGVSGHGYELKDESLKDFNMFYHYISK 845
Db 781 MPHSATARNLPENNETGLENVINKVATPKPGVSGHGYELKDESLKDFNMFYHYISK 840
QY 846 TQHSKAEHMOKKRRKQENKDEALPPPPPECFAPSKVINLLNCDIMMYTILRTVFERAID 905
Db 841 TQHSKAEHMOKKRRKQENKDEALPPPPPECFAPSKVINLLNCDIMMYTILRTVFERAID 900
QY 906 TDSNLTEGMLQMAFHIALGLLEEQQLQKAPEEVTFDYHKASRLGSSAM---NIQM 962
Db 901 TESNLTEGMLQMAFHIALGLLEEQQLQKAPEEVTFDYHKASRLGSSAMNAQIQM 960
QY 963 LLEKLKIPOLEGOKDMITWILQFDTVTKRLREKSLIVATTSGSESIKNDIEITHDKEA 1022
Db 961 LLEKLKIPOLEGOKDMITWILQFDTVTKRLREKSLIVATTSGSESIKNDIEITHDKEA 1020
QY 1023 ERKKAFAARHLROKTAQMSALOKNFIETHKLMYDNTSEMPGKEDSIMBEESTPAYSDY 1082
Db 1021 ERKKAFAARHLROKTAQMSALOKNFIETHKLMYDNTSEMPGKEDSIMBEESTPAYSEA 1080
QY 1083 SRIALGPKRGPSYTEKEVLTCLQEBQEVKIENNAMVLSACVQKSTALTQHRGKPTELS 1142
Db 1081 SRIALGPKRGPAVTEKEVLTCLQEBQEVKIENNAMVLSACVQKSTALTQHRGKPDVHL 1140
QY 1143 GEALDPLFMDPLDLAGYTYTSGCHVMHACVQKYFEAVQLSSQRIHVDLFDLESGEYLC 1202
Db 1141 GETLDPLFMDPLDLAGYTYTSGCHVMHACVQKYFEAVQLSSQRIHVDLFDLESGEYLC 1200
QY 1203 PLCKSLCNTVPIPILOPQKINSFNADALAQLLTLASWIGOTVLARISGYNIRHAKGENP- 1261
Db 1201 PLCKSLCNTVPIPILOPQKINSFNADALAQLLTLASWIGOTVLARISGYNIRHAKGAPA 1260
QY 1262 IPIFFNGMGDSITLPHSILSFGVSESIKYSNISKEMVILFATYIRIGLKVPPDERDPR 1321
Db 1261 VPVLFNGMGDSITLPHSILSFGVSESIKYSNISKEMVILFATYIRIGLKVPPDERDPR 1320
QY 1322 VPMLTWSTCAFTTOAIENTLGDGCKPLFGALQNRQHNGLKALMQFAVAQRITCQVLIQK 1381
Db 1321 VPMWTWSTCAFTTOAIENTLGDGCKPLFGALQNRQHNGLKALMQFAVAQRATCQVLIHK 1380
QY 1382 HLVRLLSVLPNLSKEDTPLLSTDLPHVLGAVLAFPSLYWDDPVDLPQSSVSSYNHL 1441
Db 1381 HLRLLSVLPNLSKEDTPLLSTDLPHVLGAVLAFPSLYWDDPVDLPQSSVSSYNHL 1440
QY 1442 YLFHLITMAHMLQILLTLVDR---GLPLAQVQDESEEAHSAFPAISQVTSIGCDI 1497
Db 1441 YLFHLITMAHMLQILLTLVDR---GLPLAQVQDESEEAHSAFPAISQVTSIGCDI 1500
QY 1498 PGWYLVWSLKNGITPYLRCAALFFHYLLGVTPPELHTNSAEGEYSALCSYLSLPTNLF 1557

Db 1501 PGWYLVWSLKNGITPYLRCAALFFHYLLGVTPPELHTNSAEGEYSALCSYLSLPTNLF 1560
QY 1558 LFOEYWDTVRPLQRRCADPALLNCLKOKNTVVYRPRKRNLSIELPDDYSCLLNQASHFR 1617
Db 1561 LFOEYWDTVRPLQRRCADPALLNCLKOKNTVVYRPRKRNLSIELPDDYSCLLNQASHFR 1620
QY 1618 CPRSADDERKHPVLCFCGAILCSQNICQEIIVGEEVACGACIFHALHC----- 1665
Db 1621 CPRSADDERKHPVLCFCGAILCSQNICQEIIVGEEVACGACIFHALHC----- 1680
QY 1666 -----KARGCAYPAPYLDEYGETDPLKRGKGNPLHLSRERYRKLHLVWQOHCIEETA 1717
Db 1681 CRVVLVEGKARGCAYPAPYLDEYGETDPLKRGKGNPLHLSRERYRKLHLVWQOHCIEETA 1740
QY 1718 RSQETNOMLFGFNWOLL 1734
Db 1741 RSQETNOMLFGFNWOLL 1757

RESULT 2
T25604
hypothetical protein C32E8.11 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T25604
R:Gattung, S.
submitted to the EMBL Data Library, February 1997
A:Description: The sequence of C. elegans cosmid C32E8.
A:Reference number: Z20056
A:Accession: T25604
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1927 <GAT>
A:Cross-references: EMBL:U08308; PIDN:AB42328.1; GSPDB:GN00019; CESP:C32E8.11
A:Experimental source: strain Bristol N2; clone C32E8
C:Genetics:
A:Gene: CESP:C32E8.11
A:Map position: 1
A:Introns: 56/3; 157/3; 255/1; 469/1; 578/3; 832/3; 907/2; 1165/2; 1218/2; 1275/2; 14

Query Match 20.6%; Score 1886.5; DB 2; Length 1927;
Best Local Similarity 27.3%; Pred. No. 2.4e-108;
Matches 553; Conservative 305; Mismatches 656; Indels 513; Gaps 67;

QY 102 QLCGRVFKSGTETYSRCDADPTCVLCMDCFPDQSVHKNHRYKMTSTGGGFCDCGDA 161
Db 14 QICGHVFNKGLTYTCLDCAITDGTVCMLQCEVSIHSHKHYKMHSSSGYCDGADA 73
QY 162 WKTGPFVYNHEPG-----RAGTIKENSRCPLNEEVIVQARKIPSPVIKYVEMTWEE 214
Db 74 WTEGYACANHEBKDDDEAAVLAPELKK--RCEQLVEIILQ---FSLSMITHKDDLPEI 128
QY 215 EKELPPELOIREKNERYCYVLFNDEHSHYDHYIYSIORALDCELAQLHTTAIDKEGR 274
Db 129 FEKKMPE--VYNEAQOYLTVLYNDETHYVESVLELYICTKQDQMLVATIVDREGS 186
QY 275 AVKAGAYAAOCBAKEDIKSHS-----ENVSOH--PLHVEVLHSHIMAHQFALRLGSW 325
Db 187 AVKLGSKADCTYKAKDDQVQKTARDPTSIRSSNNHLPLSVKVMYDITLQALQNFISILLTW 246
QY 326 MNKIMSYSDFRQIFCQACL----- 345
Db 247 LNTQMDVPPLREIVGEILLSSKFALKKNTYRKMSQDQRLVAGIIRNVWLPDDDEEL 306
QY 346 -----REEP-----DSENPCLI 357
Db 307 FALDGRMDVDEMDDDDIGEALQMLQMDADDEEITAALAGVSEHQSGPGRSDSFTTFL 366
QY 358 SRLMLWDAKLYGARKILHELHLSFFMEMEKYKLFAMEFVYKYKOLQKOEYISDDHRSI 417
Db 367 ENILLQDTQMWRKAGRSIHLQMLMRTVMFYIDQKRFKAFKMLHYNEIYEDFIKDDHMDV 426

Db 217 ESTLQDEKTSRLSENKYGIDDDSCNMYSLVLWDEKHSFKQFYEQIITALL--ELPNNVFG 274
Qy 262 QLHTTADTKGRRRAKAGAYAAOQAKEDIK---SHSENVSOQHLPHVEVLHSEIMAHOKF 318
Db 275 KMANIINDIGR-----ACTVETNIKELKIGQKLAQINLAIVSIRMRDIFREES 325
Qy 319 ALRLGSMNMKIMSYS-----SDRQIFCOACLR----- 346
Db 326 CAVLEWADIAGSSICGKKNRYFSSVICKELVRPNWNCGLHNSDUTFRLSRLALPEIVA 385
Qy 347 -EEDP-----SENPCILSRM----- 361
Db 386 IDSPDIFLNEDHINSSGSDTSSSHLETDSSIHRSRHWYPSNSLPDVLASYASRVRFDFYF 445
Qy 362 LMDAKYKGARKILHELIFSSFFMEMEYKKLFAMEFYKQYKQLOKEIYISDDHDSISITA 421
Db 446 LYDLKWKSLRYKLQELYLGVFTQPGFKEIMGARIAISYRRLAELFLLDREPHSVIF 505
Qy 422 LSVQMFVPTLARHLIEQNVISVITILEVL-----PEYLD-----RNNKFNQGY 469
Db 506 FMSQIFTVADVAKLLVTEYDFLTINATLYTFFTYKKLNTPNYVDQHAMIITDASAAPH-- 563
Qy 470 SQDKLGRVAVICDKYLILSKPTIMTERLR--MOFLEGRFSFKILTCMOGMEIIRVOV 527
Db 564 -----SRYTHIFHHIQFMISIPCV-AEIVREDLKLQYADFFNL-----FGCMCPYTRAV 614
Qy 528 GOHIEVDPD-WEAAIAIOMQL-----KNILMFQEW-----CACDEELLVAYKECHKAV 576
Db 615 SQHVENWDSMWYLVNLSQVAKLCHVGNVFMELNKNLANINYLSILILYPKARNES 674
Qy 577 MRCTSTISSKTVQSCGHSLETKSYRVSDLSIHLPLSRTLAGHLVRLSRLGAVSRL 636
Db 675 W-TNTESLTGTITVDER-GNS-KLIEYDIALQPVSFHPLHLLVYL-----L 719
Qy 637 HEFVSFDF-----QVFLV--VEPLRCLVLVAOVAEMRRNGLSLSQVFFYQDYVKCR 689
Db 720 SFYVERDNYKLLWTQDOLLAVTDPLRCVCAWLSQMRKALWTRNGTITLDQAHHRNLSFH 779
Qy 690 EEMVDKDIIMLQIGASLMDPNKFLLLVLQRYELAEAF--NKTIKQDQDLKQYNTLIEE 747
Db 780 EYTFDLDVLLQLTLTYGDDAILPSISRFQEDQMYGREFVPHKYD-VSQVTIMEE 838
Qy 748 MLQVLIYIVERYVPGVGNVTKEEVTWREIHLICIEBPMPSAIAKNLPENENNETGLEN 807
Db 839 FLILLISIVCNTAVLDHWDITR--IEYGAHILCFRPLPYSEITKTRCEHLLHOKPES 896
Qy 808 VINKVATFKK-PGVSGHGVYELKDESKDFNMYYHYHYSKTHSKAEHMOKKR-RKOENKD 865
Db 897 TLKVAFTRNAEINGDSGFTLKDEYFDYVDPFNIHYSRNQREAEANILRRRYSKQHSKH 956
Qy 866 EALPPPPPPFCPAFSK--VINLLNCIDIMYIL-RTVFERAIDTDSNLWTEGMLQMAPHI 922
Db 957 --LESVYVEYHPIHLSNITIPILQDSFVGLMHTIYVAYIYIPYDQKLEGLVNTALHA 1014
Qy 923 LALGLLEKQOLQKAPAEVEFTDFYHKASRLGSSAMNIQMLEKLGIPQLEGQD----- 978
Db 1015 CLLVLMSEK-----GSEPIFSKKICE--NRFPVVEGLQYCNYS 1050
Qy 979 ----MITWILQK-----FDIVKRLREKSLIV-----ATTSGSSIKNDE 1014
Db 1051 PDVTLFVLQCMKNHRNFVYVK-----EKISLIMKILKSEVPILLYBPVVAETLSISSKTIQ 1107
Qy 1015 ITHDKKAEARK-RKAEARLHROKIMQMSALOKNFTETHKLMYDNTSEMPGKEDSIMEE 1073
Db 1108 SLSDAEQOQHLAKVRWAKERQARIMEQFRQONKNFLENHALFASDCM-----DEADEF 1163
Qy 1074 ESTPAVSYSRIALGPKRGPSVTEKEVLTCLCOEQEVEKVENAMVLSACVQKSTALT- 1132
Db 1164 SVTSSVS--TKFLDLP-----PIDTCLLCQOELKDKRPYGLTFV---VLRSVLRL 1209
Qy 1133 -----QHRGKPIELSG-----EALDPL----- 1149
Db 1210 FPADDANYSEVLDIPDSLOHEIOERPFGLAGKRKKVLDSTPEAYDYDNYVYKKGKGNELHQ 1269

Qy 1150 -----FMDPDLAYGTYTSCGHVMAVHCQYFEAVQLSSQ---QRTHVDLFDLESGE 1199
Db 1270 LKDSFNGFPDQDRGLHATGCGHPMHIDCFKNHIAVTVTLATRANPYRNHP--HNLSMKE 1327
Qy 1200 YLCPLCKSLCUNTIPIPILOPKINSINADALAQLLTARWI-----QTVL-- 1245
Db 1328 FLUCPLCKALCUNTIPIPILOPKINSINADALAQLLTARWI-----QTVL-- 1245
Qy 1345 GKPLGALONROHNGLKALMQFAVAQ-----RITCPOVLI-----QKHLVRL 1387
Db 1489 FIPVNGKL-----HNVVK-LFSYSLQVEASTRGRHIKCSSIPADIWVHNLGKNQOVFLRL 1543
Qy 1388 SVVLNPKISEDPCLLSIDLPHVLVGAFLAF-----PSLYWDDPV-- 1427
Db 1544 S-----ESIKTYTLLC--AHDQKRIGSGIOEPFISFCQOKRIFGRLLPSL--DSPYTKSI 1596
Qy 1428 -----DLQPSVS-----SSYNHL-YLFHLITMAHMLQIILLTVDTLPL 1465
Db 1597 TDDRVEPLLVKDTREFEAEASVGLLSCDESPHYLTQLYVTADIVRNLTLLSQRNSL-L 1655
Qy 1466 AQVQEDSEA-----HSASFFAEISQYTSIGCDI-----PG--WYLWVS 1505
Db 1656 KQMSVEFEAFYEQKGEHLVQIWKSLRVDGAGLINFDCTEDDLNPHLLFTLYKL 1715
Qy 1506 LKNGITPYLRCAALFFHYLLGYTPPEELHTNSAEGYSALCSYLSLPTNLF--LFOEY 1562
Db 1716 LERFSLIFLRKCALLWYCRYGVSTQPNLNFQNSLSLQTKMHIPGVIELSNHLCILTA 1775
Qy 1563 WDTVRLPLQRRCADPALLNCLKOKNTVRYPRKR-----NSLIELPDYSCLLNQASHFR 1618
Db 1776 SSTEWSLIKHWK-----NFFTETGPLCDFPRAYPGIVELSVPELQKVFELLARRC 1829
Qy 1619 PRSADDERKHPVLFCGAILCSQNTCCQEVNGEVEGACIFALHCKA-----RGC 1670
Db 1830 SKCL-TEPMEPAICLFCGKLLCFQSHCCS--FNG--IGECNLHMOQCASDIGIFLIVKCK 1884
Qy 1671 AY-----PAPYLDYGETDPLKRGPNPLHSRERY-RKLHLVMOQHCIEIAR 1718
Db 1885 ALLYNPPVGSFVSAPFLDAYGETDGLRGRSQRYSQKRYDETIVTVMWLNINGSIPYIAR 1944
Qy 1719 SOETNQMLFGFNWQLL 1734
Db 1945 QLDANPDTCG--WETL 1958

RESULT 4

T37711

probable n-end-recognizing protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T37711
R:Murphy, L.; Harris, D.; Wood, V.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, June 1997

A:Reference number: Z21738
A:Accession: T37711
A:Status: preliminary; translated from GB/EMBL/DDJ
A:Molecule type: DNA
A:Residues: 1-2052 <MUR>
A:Cross-references: EMBL:Z97208; PIDN:CAB10108.1; GSPDB:GN00066; SPDB:SPAC15A10.11
A:Experimental source: strain 972h-; cosmid c15A10
C:Genetics:
A:Gene: SPDB:SPAC15A10.11
A:Map position: 1

Query Match 9.5%; Score 866; DB 2; Length 2052;			
Best Local Similarity 20.2%; Pred. No. 5.3e-45;			
Matches 443; Conservative 320; Mismatches 724; Indels 702; Gaps 90;			
QY	67	QBSVQMSIPTPLEW-YLFGEDP-----DTCLEKLKHS-----GAFOLCGRVF	108
Db	43	QEVFLSILNEDNWKYFLKKGAIITSDPRLSRLOHSEPECAQLODKRSGSKVCGHVF	102
QY	109	KSGETYSRCDCAIDPTCVLCMDFQDSVHKHNRKYMHTSTG-GGFCDCGDTEAWKTGPF	167
Db	103	RAGEVIYRCKNCGLDNTVCAPCFHATNHEGHEHVSISTSYSGICDCGDGPEAWNDLN	162
QY	168	CVNHEPGRAGTAKENSRCPLNEEVIQVARKIFFSVIKYVVMETW-----	212
Db	163	CKTH-----NVPDDEQKPEEVI-----PLEQHSRTTHILLDFILDVFCSPV	209
QY	213	-----EERKELPELQIREKNRY-----XCVLFNDRHSHSDHYIS	249
Db	210	NLKAQSTVGSILLADEAS-----RLSSAKYGVADRPCNVFRVWLWDEVHFDAVVG	262
QY	250	LQALDCELAEOHLHT-AIDKEGRRAVKAGA-----	280
Db	263	VLEALDSSNTAFGLEVAQVRDSTGRFAVATSASVHEAIRIANAIKENLAVNVRTARDF	322
QY	281	-----YAA-----COE-----	286
Db	323	REDICGILLWFDDLLSHVCYFADYLQIIVCDEILKNWSPGLEKPAKPVNFNNLPLEI	382
QY	287	-----AKEDIKSHSNVS-----OHPL-----	303
Db	383	VNDDSEDDIYAAEELLDVIANLODETGVTRIANLGDEDFEADMTDPTTAGDHPDLD	442
QY	304	-----HVEVLHSEIMAHQF-----ALRLGSMWNKIMSY-----	333
Db	443	NDVNLLDFETEREDIDLTDEVMEETENEAADYGVNRRNTRODDVQDISMETESQNE	502
QY	334	-----SDFROIF-----	340
Db	503	TDESQNTENVYDNPQTHTPVPIPTATQDVVTRPFNSOLLNNLRQIINARRRPAAV	562
QY	341	COACLREE-----PDS-----ENPCLISRL---MLWDAKLYKGARKILHELIFSSFF	384
Db	563	CQVSLREDYKWSHPPIPPSYSFVESPSILRLDYFLFLDKFWKRLUGLLSKLYVVPFN	622
QY	385	MEMEYKFLFAMEVYKYLQKEVYISDDHRSISITALSQVMEFTVPTLARHLIEEQNVIS	444
Db	623	RNLLFRLMGIRVHYRSLATAFADREPDSHVMFLSVQFTTSPSLAAVVKDYDFLT	682
QY	445	VITETLEVL-----PEYLDNRNKNFNQYSQD-----KLGRVYAVICDLK	485
Db	683	NLNATTLSSLTQSNRPSTL-----FSSDIEVTPTIQLNRQVLKTRTYNLFSDLG	732
QY	486	YIL-----ISKPTTWELRMQFLGPRFLKILTCQGMEEIRQVQCHIEVDPDWEAAI	541
Db	733	YLLQHPQVKLVVDDTRYHQYIDLRFV-----QGVIPQORAILSHVQ-----WDFPH	781
QY	542	AIQWQLKNILLMFOEACACDEEL-----LLVAYKECHKAVMRCSTSFSSSKTYVQSCGHS	597
Db	782	G-----KNILFVMQVRAMLNTVSSCFQAPYERLFAI-KCIITSIHPKLDI-----AES	832
QY	598	LETKS-----YRVSEDLVSIHPLSLRSLAGLHVRLSRLGAVSRHLHFVS	641
Db	833	LEPLSCIPSSSLTNFTQPLVPFVSRRDPISFYHP-----LHWMLSNLFSYCRVDASH	885
QY	642	FEDQVEVLVEYPLRCLVLVAQVVAEMRRNGLSISQVYFQDVQVKCREEMYDKDITMLQ	701
Db	886	WDKDTLLALLDHLRVCVLLAQDCNLRWNGRSILLTDAFYROLNNIEVSYDKDILAIQ	945
QY	702	IGASLMDPNKFLLLVLQRYELAE-AFNKTISTQDQDLIKQYNTLIEMLQVLIYIGERY	760
Db	946	TILMFDPNVLNAVYQREFTDWLYNLTNEHPNDYTERIPAMLCMKMLLELIALITER-	1004
QY	761	VPQGVNVTKEEVTMREIHLICTEPHSAIAKNLPENENNETGLENVINKVATFKKP-G	819
Db	1005	-EQILHVDIIDIIRTLAQQLCFGLAYSALLSTISSNLVESLSPDKIREEYTSYKAPDG	1063
QY	820	VSGHGVTELKDESLKDFNMYYHYHYSKTOHSAEHMOKRRKQEN--KDBALPPPPPPPPFC	877
Db	1064	LHDFGVSLKDEYDLDVDPYFYHYNKNREESDTILKRLAKNNVSAESIIEPKIRFL	1123
QY	878	-----PAFSKVINLLNCIDIMMYILTRVFERAIDTDSNLWTGMLQM-AFHILALG--LLE	929
Db	1124	EKGDHIDFFAANASTFSLIIF-----RAIE-----YALVOAESGSSDIGNTIILG	1169
QY	930	EKOOL-----QKAPBEEVTFDFVHKA-----SRLGSSAMNIQMLLEKLKIPOL	973
Db	1170	DALQCLLSIKMHEFSKSNDFCSRCAERYPTDSSIMREFGSAVCLAECLFAILKSPKY	1229
QY	974	EQOKDMITWI---LQMPD--TVKRLREKSLIVATTSSGSEIKNDEITHDKKAERKRA	1028
Db	1230	KDVHVKNVAVLAGLQKNDPSAYSNNLEATHFELSTSS-----TSDSNEIEKTQEK	1280
QY	1029	EAARLHRO-KTMAQMSALOKNFIETHKLMYNTSEMPGK---EDSIMEEESTPAYSIDYR	1084
Db	1281	KRLALEKOKKIMQOFRDQOASFLA-----QNTDFDIDGQDTEDEVTEEPVEEKYHEH	1334
QY	1085	IALGPKRGPSYTEKEVLTCILCOEDEVK-----TENNAMVLSACVQKSTALTQHRG	1136
Db	1335	I-----RG-----NCLLCQECNDQAPYGVIGIIOGSSLLRKTDVHSEIILDEIYS	1380
QY	1137	KPIELS-----GEALDPLFMDPD-----LAY-----GTYTSGCGHVMHVCWKQY	1176
Db	1381	VPPNLDRESHSRPFKKYDVTVFNRSKDRLLSAYPGNNIRGVFVSGCGHMLHLCFKNY	1440
QY	1177	FEAVOLSSQQRHVDLFDLESG--EY-----LCPCLSKLONTVPIIPILOPOKTN	1224
Db	1441	YVARSMYRN-----DVTAGLSYKYYKSTAKFFMCPCLRSLSNVLLP-MPQIPKM--	1489
QY	1225	SENADALQALTLARWIQTVLARISYINIRHAKGENPIPIFNQGMGDSITLFSHSI---	1280
Db	1490	CLNIDTLNPRSMNGWLEEI-GTMSSSFQEQ-----LVRSLSDDTKDTRFSCFLRP	1540
QY	1281	-----LSFGVSSISIKYNS--IKEMVILFAITIVIRIGLKVDPDRDPRV-PM	1324
Db	1541	WINSKIISAMLARUKIADGALIDQNNRDVSDLYDRYCETT-KLANKLVKGSTFNVSPH	1599
QY	1325	LTWSTCAFTIQAENLGLDEGP-LFGALQNRQHNGLKAL-MQFAVAQRITCPQVLIQKH	1382
Db	1600	DLNLSAVTVSSLE--VSQRCKPKQSGATRSWFNENLGLTSLFPLTSLD-----	1648
QY	1383	LVRLLSVVLPNIKSDPTCLLSIDLFLHVLGAVLAFPSLY-----WDDPVDLOPSSV	1434
Db	1649	---VLKVCDDQIIRKSDQOALLMESQKLLVCKIFYRHSQKMLNRMGRMSDHDQOPFL	1705
QY	1435	SSSY-----NHLYLFHLITMAHMLQILLTVDTGLPLAQOED---SEE	1474
Db	1706	SNTDFDFVKISSMLIFGQDNILYVKLFYLSSEICKTIISMIVVADSSVVDPDLTINTS	1765
QY	1475	AHSASSFAEISQVTSIGSGICDIPGWYLWVSLKNGI-----TPYL	1514
Db	1766	QOOSQOYF-----ILC--KNVLLWSSNNIEILDDSNLLRLMSLYEKSLSPL	1813
QY	1515	RCAALFHYLLGVTPP--EELHTNSAEGEYSALCSYLSLPTNLFLFQEWY-----DT	1565
Db	1814	RRVALVLYCMFDISLEFNEFSNEDDSELERLSKLKIVPP-----LQELYSQMSSDENQ	1868
QY	1566	VRPLLQRRCADPALLNCLKQKNT-----VVRYPRKNSLIE-LPDDYSCL	1609
Db	1869	ILELIAGWCEHLA-----QNTWGDSTISLEPGIYELVKLPHELNIDMSQMSVCCM	1921
QY	1610	LNQASHFRCPRSADDERKHPVLCFCGAILC---SONCCQEQIVNGEEVGACIFHALHC-	1665
Db	1922	CHKTPIL-----PAICMLCGSVICFNARQNTVSSRRLTGE-----CNKHAATCT	1965
QY	1666	-----KA-----RGCAYPAPYLDEYGETDPLGKRNPLHLSRERYR-KLH	1704

Db 1966 GSVGIFITKACGILLDLSISNTGTIMPTPYLDIHGETDLQLRRCGPOFLNQKRYDFVVR 2025
Qy 1705 LVWQOHCIIIEIARSQETNQMLFGFNWOL 1733
Db 2026 EQWLROTVALQKMARHMDWTEMQ---NWRM 2051

RESULT 5

Sl2332
ubiquitin-protein ligase (EC 6.3.2.19) - Yeast (Saccharomyces cerevisiae) (strain S288C)
N:Alternate names: protein G7168; protein YGR184c

C:Species: Saccharomyces cerevisiae
C:Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 03-Jun-2002

C:Accession: Sl2332; S64502; S64498

R:Bartel, B.; Wuening, I.; Varshavsky, A.

EMBO J. 9, 3179-3189, 1990

A:Title: The recognition component of the N-end rule pathway.

A:Reference number: Sl2332; MUID:91006011; PMID:2209542

A:Accession: Sl2332

A:Molecule type: DNA

A:Residues: 1-1950 <BAR>

A:Cross-references: EMBL:X53747; NID:g4743; PIDN:CAA37779.1; PID:g4744

A:Experimental source: strain S288C

R:Arroyo, J.; Garcia-Conzalez, M.; Garcia-Saez, M.I.; Sanchez-Perez, M.; Nombela, C.

submitted to the Protein Sequence Database, May 1996

A:Reference number: S64499

A:Accession: S64502

A:Molecule type: DNA

A:Residues: 1-1950 <ARR>

A:Cross-references: EMBL:Z72969; NID:g1323325; PIDN:CAA97210.1; PID:g1323326; MIPS:YGR184c

A:Experimental source: strain S288C

R:Hebling, U.; Hofmann, B.; Delius, H.

submitted to the Protein Sequence Database, May 1996

A:Reference number: S64003

A:Accession: S64498

A:Molecule type: DNA

A:Residues: 1615-1950 <HEB>

A:Cross-references: EMBL:Z72969; MIPS:YGR184c

A:Experimental source: strain S288C

C:Genetics:

A:Gene: SGD:UBR1; PTR1

A:Cross-references: SGD:S0003416; MIPS:YGR184c

A:Map position: 7R

C:Superfamily: ubiquitin-protein ligase

C:Keywords: ligase; protein degradation

Query Match 7.6%; Score 691.5; DB 2; Length 1950;

Best Local Similarity 20.0%; Pred. No. 3.3e-34;

Matches 409; Conservative 321; Mismatches 734; Indels 581; Gaps 86;

Qy 12 GGTERTMISBELPQTTPORLASWDDQVDFYAF-LHHLAQLVPEYFAE-----MDPDL- 64

Db 33 GPTERADMSRALKEFIYR-----YLYFVLSNGENLPTLFNAHPKQKLSNPETL 81

Qy 65 ---EKQESVQMSITPLEWYLFGE-DPDIKLEKLSHGFQPCGGRVFKSGTTCYSCDC 120

Db 82 VFPDSLEDAVIDKITSQOTIPFYKIDESRIGDVHKHTG--RNCGRKFKIGEPLRYCHEC 139

Qy 121 AIDPTCLVCMDFQDSVHKHURY--KMTSTGGGFCDCGDTAEWKTPGFC-----VN 170

Db 140 GCDTCLVICIRFPKDVHNVHVCDDICTETSTGICDCGDEAMNSPLHCKAEQENDIS 199

Qy 171 HEPGRAGTIKNSRCPNNEEV-IVQARKIFSVIKYVEMTWEEKELELPPELQ----- 223

Db 200 EDPATNADIKBED--VWNSDVNIALVELVLAEVDFYID--VFNONIEPLPTIQDKITK 255

Qy 224 IREKN-----ERYQVL 235

Db 256 LREMTQOQKMYERAQFLNDLKYENDYMFDTGTTAKTSPNSPEASPLAKIDPENYVII 315

Qy 236 FNDEHSHSDHYIYSQALDCELAQHLTAIDKEGRRAVKA-----GAYAACOE- 286

Db 316 YNDEYHNSQATTALRQGV-PDNVHIDLTLSTRIDGEGRAMLKCSQDLSSVLGGFFAVQTN 374

Qy 287 -AKEDIKSHSENVSOHPLHVEVL---HSEIMAHQKFAALGLGSMNMKIMYSDDF----- 336
Db 375 GLSATLTSWSEYLHOETCKYIILWITHCLNIPNSPQTFRNMGMKTL--CSEYLNATEC 432
Qy 337 -----RQIFCOACLRREP-----DSEN 353
Db 433 RDMTPVVEKYFNKFKNDPYRYIDLSILADGNQIPLGHHKILPESTHSLSPLINDVET 492
Qy 354 P-----CLISRLMLDAKLGKARKILHELIFSSFFMEMEYKKLFAEFVKYKQLOK 406
Db 493 PISRTSYNTRLOHILYFONRYWKLRKDQIONVIPTLASSNLKYKPFQCOVVEIFNHIR 552
Qy 407 E---YISDDHRSISITALS---VOMFTVPTLARHLIEEONVISVITETLEVLPEYL--- 458
Db 553 SVAYM---DREPOLTAIRECVQVLTCTPCTNAKNIPENQSFQDIW--SIIDIFKECKVE 607
Qy 459 -----DRNNKFN-OGYSODKLGRYVAVICDLKYILISK---PTIWERLRMQFEGFR 508
Db 608 GGVLIQRVQKSNLTKSYSISFKQGLYTV-----ETLLSKVHDPI--PLRPKEI---I 656
Qy 509 SPLKILTCMQGMEERQVQHI-EVDPDWEAAIAIQMOLKNILLMFQEWG-----ACDE 562
Db 657 SLLTLCKLFGANGKIKRKEGEHVHLEDQNFISVLEVTTSIYSIQIAEKVSEKSKIDS 716
Qy 563 ELLLVAYKECHKAVMRCSTSFSSSKTVVQSCGHSLETKSYRVSDELVSIIHLPLSLTAG 622
Db 717 KLFNA-----IRIISFLGNRSYLYKLYDSHEVIFKFSYSHERVAFMNPLOTMLSF 768
Qy 623 LHVRLSRLCAVSRHL---EFVSFEDQVEVVEYPLRCLVLVAQVVAEMWRNCLSLISQ 679
Db 769 LIEKVSLLDAYEALEDCSDFLKISDFS-----LRSVVLCSQIDVGFWRMGSVLHQ 820
Qy 680 VFYYQDVKCREM--YDKDIIMLOIGA--SLMDPNKFLLLVLQRYELAFANFKTISTKQ 735
Db 821 ASYI---KNPELGSYSRDIHLNQLALWERDDIPRIIYILDRWELLDFWFTGEVDYQHT 877
Qy 736 DLIKQNTLIEEMLOVLIYIVGERVYVGVNVTKE---EVTREIILHLICIEPMHSAIA 792
Db 878 VYEDKISFTIQFIQFIQIYQILTERQYFKTFFSLKDRRMDQIKNSIYLYMVKPLSYSKLL 937
Qy 793 KNLPE-ENNETGLENVINKVATFKP-GVSGHGVYELKDESKDNMFFYHYSKTQHSK 850
Db 938 RSPDYLTEDTTFDEALEEVSFVEPKGLADNGVEKFL-ASL-----YAKVDPLK 987
Qy 851 AEHMK-----KRRQENKDEALPPPPPEFCFAFSKVINLNCDIM---MYLRT 898
Db 988 LNLNEFFESSATIKSHLAKDKDEAKVVLIPQVS-----IKQDKDALNLGATRNT 1041
Qy 899 VFERAI-----DTDSNLWTEGMQLMAFHILALG-LLEEKOOLQA-----PEEVTFD 945
Db 1042 VFAKVYKLLQVCLDMEDSTFLNELHLVHGIFRDDDELINGKDSIPEAYLSKPCICNLLS 1101
Qy 946 FYHKASRLGSSAM--NIOMLEKLIKIPQLEGOKDMITWILQFMDVVRKLRKESCLIVAT 1003
Db 1102 IANAKSDVSESIVRKADYLLEKM-----IMKKPELNFESLIA- 1139
Qy 1004 TSGSESIKNDIETHDKEK-----AERKKAERAAHLRQKIMAOMSALOKNFIEH 1053
Db 1140 -----SFGNOYVNDYKDKKLRGVNLOETEKERKRLAKHQAARLLAKFNNOQTQFMKEH 1194
Qy 1054 KLMYDNTSMPGKEDSIMEESTPAVSQYSDYRIALGPKRGPVTEKEVLTTCILQEQEDEVK 1113
Db 1195 ESEFD---EQDNDVDWV-----GEKVESEDFTCALQDSS--- 1227
Qy 1114 IENNAVLVSACVQKSTALTQHRKPIELSGEALDP-----LFMPDPLA 1156
Db 1228 -STDFEVIPA-----YHDSPIFRPGNIENPEFMPMDGFFYNDEKQAVIDDVL 1277
Qy 1157 YG-TYTGSGC-----HVMHVCWKQYFEAVQLSSQORIHVDLFDLESGEYCLPCK 1206
Db 1278 EALKENGSCGSRKVFVSCNHHIHNCFKRYVQKKRFS-----SNAPICPLCQ 1324

```
QY 1207 SLCTVTIPILOPKINSENADALACOLLTLARWIOTVLARISYINIRHAKGENPIPIFF 1266
Db 1325 TFSNCTPLP--CQTSKANTG-----LSDFPLESELSL----- 1355
QY 1267 NOGMGSTLEFHSILSFGVSESIKYSIKEMVILFATTIYRIGLKVPPDERPVPMLT 1326
Db 1356 -----DTLS--RLFKPFTBENTYNTINSIFSLMI----- 1381
QY 1327 WSTCAITQAIENLLGDEGKPLFGALONROHNGLKALMQFAVAQRITCPOVLI-----Q 1380
Db 1382 -SOCQFDRKAVRKRAFNHSHKDVSLILSVHWANTISML---EIASRLEKPKYSISFFRSREQ 1437
QY 1381 KHLVRLLSVVLNPKIKSEDPCLLSIDLFLVLGAVLAFPPDLQVDPDLQSPSSVSSYNH 1440
Db 1438 KY-----KTLKNI-----LVCIMLFTFVIGK-----PSMEF-EPYPQPDVWNO--NQ 1478
QY 1441 LY-----LFHLITWAHMIQIILLTVDTGLPLAQOVEDSEAHASGSAFFAEISQ----- 1487
Db 1479 LFQYIVRSALFSPVSLRQTVTEALTTFSRQFLDFLQGLSDAEQVTKLYAKASKIGDVLK 1538
QY 1488 -----YTSIGCDIPGWLVLWLSLKNGITPYLRCAALF---PHYLL 1525
Db 1539 VSEQMLFALRTISDVRMEGLDSEIYDCLAYTELLKSL-----LPTIRRCVLFTKVLHEL 1594
QY 1526 GVTPEELHTNSAEGYSALCSYLSLPTNLFLFOEVDWTVRPLQORCADPALLNCLKQ 1585
Db 1595 KUSENETLVINGHEVEE-----LEFETAEPVNAKMKITEKESLDVLLTT 1641
QY 1586 KNTVRYPRKRN-----SLIELPDDYSCLLNQASHPCPRSD-----DERKHPV 1630
Db 1642 QESIVSHPYLENIPYECGIKLDLSKYLVNTVTSQKEIKLEERSOHNKMNADNRDLFK 1701
QY 1631 LCLFCGAIL-----CSQNTCCQEIYNGEVEGACIFHALHCKARGCAYPA 1674
Db 1702 ICILTCGVKHLRADRHMTKHLNKNCKPFGAELMPNSSEV--CLH--LTQPPSNIFISA 1757
QY 1675 PYLDEYGEYD--PCLKRGNPLHLSRERYKHLHLWQOHCILEETARSQE-----TNQM 1725
Db 1758 PYLNSHGEVGRNARMRGDLTTLNLRKYEHLNRLWINNEIPGYISRYMGDFRVTILSNFG 1817
QY 1726 LFGFN 1730
Db 1818 LFAFN 1822

RESULT 6
T30554
ubiquitin-protein ligase E3 - yeast (Kluyveromyces marxianus var. lactis)
C:Species: Kluyveromyces marxianus var. lactis, Candida sphaerica
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jan-2000
C:Accession: T30554
R:Waller, P.R.H.; Varshavsky, A.
submitted to the EMBL Data Library, April 1998
A:Description: Kluyveromyces lactis UBRI, the recognition component of the N-end rule pa
A:Reference number: Z20856
A:Accession: T30554
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1941 <WAL>
A:Cross-references: EMBL:AF061554; NID:g3114670; PID:g3114671; PIDN:AAC15841.1
C:Genetics:
A:Gene: UBRI
C:Superfamily: ubiquitin-protein ligase

Query Match 6.9%; Score 628.5; DB 2; Length 1941;
Best Local Similarity 20.4%; Pred. No. 2.7e-30;
Matches 406; Conservative 315; Mismatches 683; Indels 587; Gaps 92;

QY 40 FYTAFUHLAQLVPEIYFAEMDPDLKQESVQMSIFTPLWYLFEGEDPDICLEKLKHS 99
Db 51 YYYMISDSGRLLPLPHMTATNREFFKNVDQAMEIKLSKPKYKIDENG--HSPKFNHAG 108
QY 100 AFQLGRVFKSGETTYSCRDAIDPTCVLCMDCFQDSVHKHRY--KMHTSTGGGFCDCG 157
```

```
Db 109 --RICGAKFRVGEPIYRCCKECSFDDTCVLCVNCFNPKDHHVGHVVTSTCTEENNGICDCG 166
QY 158 DTEAWKTGPPFCVNHPEGRAGTIKENS-----CPLNEEVIYQARKIFPFSVIKY 205
Db 167 DKEAW-----NHENCKGA--EDNGRLEDEFDHDKGKISKLMESVLIELDFHIDVFNQ 218
QY 206 VVE--MTIWE-----EERELPPELOIREK-----NER----- 230
Db 219 NIEPLTTIQKPLIAKLRYFNPERIYEQADMLRLRAYNQYNDSEESNKRHLTSLDPLS 278
QY 231 ---YCVLFNDEHSHVDHVIYSIQRALDCELAEAQLHTAIDKGRRAVKAGA--YAAQC 285
Db 279 TLKDYAILVYDFEHNYSQASAAIRQG--GPDNKHIDLLTAKIDSGRSLRCSADIASLM 337
QY 286 EAKEDIKSH--SENVSOHPLHVEVLHSEIMAHQFALRGLSGMMKIMSY--SSDFRQIFQC 342
Db 338 GRIFSQSNGLSCTITQ--WVEYLHQEAC--KYSI---MWINDCLNIPNSTFQSLFRN 388
QY 343 A-----CLREEPD-----SENPCLI----- 357
Db 389 AIGKVLCSKVEPFYQSIDMTSVVRDYFSDSYLSDDPYLYADHSVLGEGVKIPLGRHKSUD 448
QY 358 -----SRL---MLWDAKLYGARKILHELIFSEFFEMEMEVK 390
Db 449 PGDISAISPILUNKVIAEDHHEYTNSRQYVLFLENRYWKKLRKIVQDILLIPLTASSAVOK 508
QY 391 KLFAMEFVKYKOLKOKEYISDDHDSISITALS---VOMFTVPTTLARHLIEE-----Q 440
Db 509 PMFTDQLVEIPPHMTRS--GTFMDREPOLTSRESVQVOLFCTPTATSYIFHSGHENYLIW 566
QY 441 NVISVITE--TLLE--VLPEYLDNRNKNFQGYSDKLGRRVAVTCDLKYI-----LISKP 492
Db 567 SVIDFVDFSTMDGTLVWQVRQSRNP--SKSYISISFKQGLYAVETLLSKITDPNLLPK 624
QY 493 TIWTERLRMOFLEGFRSFLKLT---CMQGMEEI--RRQVGQHI--EVDPDWEAAIAIQMOL 547
Db 625 -----GEFIMITVLCKFNAGWKTKKREGEHVLREDQHFIPLEYBTTSV 669
QY 548 KNILLMF-----QEWACADELLLVAYKECHKAVMCRSTFSSSKTVVQSGCH----- 596
Db 670 YSIIQTEDKVLQOSKDHIDQRLIGA-----INLDSFL-----GHRNLSVK 711
QY 597 ---SLETKSYRVEDLSYIHLPLSRTLAGL--HVRLS--RLGAVSRKLFHEVDFEDQVEYL 650
Db 712 LYKDFEIIKFOISKEQVSFMPNVTFLCFVQHVPLQVSIQVLSQSKDYLVISDF----- 766
QY 651 VEYPLRCLVLVAQVVAEMWRNGLSLISOVFIYQDVCKREM--YDKDIIMLQIGASLMD 708
Db 767 ---ALRSVVLCSQIDIGFWNRGMSVLHQSAIY---KNPMSYSSYRDIQLNQL--AFLIE 819
QY 709 PNKFLLLV---LQRYELAAEFNKTISTKDQDLIKQYNTLIEMLQVLIYVIGERYVPGVG 765
Db 820 KNDQRYIYNMLDRLWELLDWDFGVPSTETVYDDKISSIIQOFVAFVLQILVER----- 873
QY 766 NVTKEEVTMRB-----IHLICIEPMPHSAIAKNLPEN--ENNETGLENVINKVATF 815
Db 874 DFYKKFDTLEETQLYNIKNALIIYKLYAEPLSVTDLNDIPDYLTESVSQFDTVLEVSIV 933
QY 816 KKPQVSGHGYELKDESUKD-----FNM--YFHYKSTQHSKAEHMKRKRQKNKDEA 867
Db 934 IEPKGSIKWCFQTKEGTLQENRFLRLNMGNDFEHSATIVKSHLADSKEKRAKIIIVKPQ- 992
QY 868 LPPPPPPFCFPAFSKVINLNCIDIMMYILTRVFERAIDTNSLDNLTWTEGMQAFHILALGL 927
Db 993 -----LLEDELDPGAR---ELGSTRNLFA---KLIFKULKLAV 1027
QY 928 LEEKQQLKAPEEVETDFYHKASRLGSSAMNIQMLLEKLGIPOLEGQKDMITWLOMF 987
Db 1028 -----SDSSSFYTELHLHLIAIFRDEDMVNGKOSLPAYISKPICDLLSIV 1075
QY 988 DTVKRLREKSLIIVATTS-----GSEISKNDIEITH-----DKE 1020
Db 1028 -----SDSSSFYTELHLHLIAIFRDEDMVNGKOSLPAYISKPICDLLSIV 1075
QY 988 DTVKRLREKSLIIVATTS-----GSEISKNDIEITH-----DKE 1020
Db 1028 -----SDSSSFYTELHLHLIAIFRDEDMVNGKOSLPAYISKPICDLLSIV 1075
```

Db 1076 DSEGSFSEN--VVATADYLLDNIMKRPTAVLESLETCFCTKYIADYKIRKANQGVNFE 1133
QY 1021 KAERKRAEARLHRQKIMQMSALQKNFIETHKLMYDNTSEMPCKEDSIMEESTPAVS 1080
Db 1134 ETEQERKRLAKNRQOQIMNFRSQQKFMKDHE-EYSAGND----EDVMDGEDLAG-- 1186
QY 1081 DYSRIALGPKRGPSTKEVLTICLCOEQEVKTEENAMVLSACVOKSTALTQHRGKPIE 1140
Db 1187 -----ELNEHCSLCHDD-----VSDFFVIPIQNTYSPVFLSNPTPME 1226
QY 1141 L-----SGEALDPLFMDPLDAYGYTG-----SCGHVMHAYCWQKYF 1177
Db 1227 IYKPMHGFDDNEHL--ATYNTDLYFKKENGASQOLMHSTOKVLVSCNHAHYRCFKHYI 1284
QY 1178 EAVQLSSQORHVDLESEGYCLPCKSCSICNTVPIPILOPKINSADALAQLTL 1237
Db 1285 D-----KKRYSTDLE-----ICPLCQTCNSVIPV-----DTVKLQSGDRLLQOKLT 1326
QY 1238 ARWQTVLARISGYNIRHAKENPIPIFFNOGMGDSITLFFHSILSFGVESSIKYSNIKE 1297
Db 1327 GGLDESULLTFSEYSSE-----CNDEVGKIIIL-----SLKDSNN-- 1360
QY 1298 MVILFAUTTIYRIGLKVPPDERPRVPMLTWTSTCAFTIQAIR-----NLIGDGGKPLF 1349
Db 1361 -----GURL--NRNDP-----TWIQDRPLTSLQFSNNICLLEMLRLNKDPF 1401
QY 1350 GALONRQHNGLKALM--QFAVAQRITCPQVLQKH-----LVRLL 1387
Db 1402 GTLLSGBEQKFTLQNLILKSLAVYTRLT-----KHEEYMSNSMKIFVAAYIQVIRFF 1454
QY 1388 SVVLPNTKSET--PCLLSIDLFLVGVAVLAFPSLY--WDDPDVLQPSVSSSYNNHLYL 1443
Db 1455 RVVETVLKSRSLSFCDLQEV-LVERLKLGTDFGSRKYESELRAQTCCLDSSEFS----- 1509
QY 1444 FHLITMAHMLQILVTDTGLPLAQVQDESEAHSSAFFAISQYTSIGCDDIPGWILW 1503
Db 1510 -----IVLKTITLGAAGF-----DQVEKHTLDLFL----- 1534
QY 1504 VSLKNGITPYLRCAALFHYLLG-VTPPEELHTNSAEYSAICLSYLSLPTNLFLLFQBY 1562
Db 1535 TFLISELLPTLRRSILLKALKQWFTGGDDLDNFNEKVDLSGLASE-SKEKHFLH----- 1589
QY 1563 WDTVRPLLQRCADPALLNC---LKQNTVVYRPRKNSLIELPD---DYSCLLNQASHF 1616
Db 1590 ---IRFLIQTDFD-LLMNSHSPSSPLSNAPHDYCSIIKLTDLATHLNTYVTNNKNI 1645
QY 1617 RCPRSADDERKHPV-----LCLFCGAILCSQNICQCEIVNGEEVGACIFHALHCK----- 1666
Db 1646 TLREENDOKIRNTVNRLDYKICLCIGVKIHAR-----TDGLEMQK---HMERCSHGSS 1695
QY 1667 -----ARGCAY-----PAPYLDVEGETD-PGLARGNPLHLSRERYKRLRLVW 1707
Db 1696 GLFLIPNISQVCLYLSRDPDCTVNISAPYLSNHSGESGRNAIERGDLTVLNHARYEHLTRLW 1755
QY 1708 QOHCIIIEIAR 1718
Db 1756 ISNGIPGVISR 1766

RESULT 7
T48251
ubiquitin-protein ligase E3-alpha-like protein - Arabidopsis thaliana
N:Alternate names: protein T4822.60
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T48251
R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft,
submitted to the Protein Sequence Database, March 2000
A:Reference number: 224489
A:Accession: T48251
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1225 <BEV>

A:Cross-references: EMBL:AL162874
A:Experimental source: cultivar Columbia; BAC clone T4E22
C:Genetics:
A:Map position: 5
A:Introns: 21/3; 390/3; 785/2; 840/3; 914/3; 1148/3
A:Note: T4E22.60
Query Match 6.6%; Score 606.5; DB 2; Length 1225;
Best Local Similarity 21.1%; Pred. No. 3.1e-29;
Matches 275; Conservative 190; Mismatches 422; Indels 417; Gaps 47;
QY 60 MDPDLKQ-EESVQMSIPTLEWYLFGEDEPDICLEKLKHSAGFA--LCGRVFKSGGTTVS 116
Db 79 VSPTKKRRFRESNM-----LOWLMFQDEPVSRLNLAKLNDQRCVCSWGQNDIAR 133
QY 117 CRDCAIDPTVCLMDCFDQSVHKNHRYKMHTSTGGGFCDCGDTAEAKTGPFCVYHPPGRA 176
Db 134 CRTCENDPTCAICVPCFQNGDHNHDSYI-IYTGCGCCDGDGTAKKPDGFCSNHKG--- 189
QY 177 GTIKENSRCPNEEVIVQARKIPPSVIKVVEMTWEEKEKLPPELQIREKNERYVCVLF 236
Db 190 ---SEQIR-PLSENLANSGVPIILDAL-----FTCWNKK----- 218
QY 237 NDEHSHYHVIYSLOALDCELAEALHTTAIDKEGRRAVKAG-AYAACQEAKEDEIKSHS 295
Db 219 -----LLSAESSQKGARSNDTLVLQKMSNELTF-- 248
QY 296 ENVSQPHLVEVLHSEIMAHOKFALRLGSMWNKIMSYSDFRQIFQOACLREBPDSNPC 355
Db 249 -----IVVEMLLFESMSSEL---LSFVSRIISSG----- 277
QY 356 LISRLMLDAKLYKGARKILHELIFSSFFEMEMEKYKLFAMFVKYKQLOKEYISDHDHR 415
Db 278 LLSILKAEERLDQDMVKMLHDL-FLKLIGDPVKECEFAKAFVYVYVISEVVKQGTDN 336
QY 416 SIS-----ITALSVQMTVPTLARHLIEQNVISVITETLLEVLPEYLDNNKNFQOGYSQ 471
Db 337 AFKKYPLLSTFSVQVLTPTLPFLVKENMLLAMLGLCLSDIFVSCSGEDGLL-----QA 391
QY 472 DKLGRV---YAVICDLKYI---LISKPTIWERLMOQFLEGFRFLKLTICMOGMEEI 523
Db 392 TKLELCETSERVIGDLKRFVMSHAIVSKYATHEHR-----ELSRWLLLLTFAQGMNPL 445
QY 524 RQVQGHIEVDPDW-----EALLVA-----EAAIAIOMQLKNILLMFOEWCACD- 561
Db 446 KRETGIPIDEENDYMHFFVLGHSIAVHSLLVNGTYSAAASDEIEIDNRNAKEEFDKCDG 505
QY 562 -----EELLVA-----YKECHKAV 576
Db 506 DGERYAKVGRLSHEDSVCTAIVSSSSFDSSMASSEVHKIDPFHALLPSSAIYLIIECLKVL 565
QY 577 MRC-----STFSISSKT----- 589
Db 566 ETCGLNDEGISKFLCKLSSSGRNIPESKMSWPRDLLNVETGGVSSNLIASSRDPSTG 625
QY 590 -----VVO-----SCGHSLETKS----- 602
Db 626 LSPLCGDIOTNLSDNWCOPYGVVQTDVTADSKRVSCNSADLTKNASGLRILGLCDWPI 685
QY 603 -YRVSIEDLVSIHPLSRLTAGLHVRSL-----GAVSRLEHFVSPEDFOVEVL--- 650
Db 686 HYDVSSQAISVHLPLHRLLSLLIQKALRICYGESASVNGVISISHE-IPHADPFSSVIGDF 744
QY 651 -----VEPRLCLVLVAQVVAEMWRNGLSLISQVFFYQDVKCR--EEMYDKDIIM 699
Db 745 HPCGFSALVMEVHLQIRVFCAQVIAGMKWKNKGSAL-----VSCSEWSEQLDLFL 796
QY 700 LOIGASLMDPNKFLLLVQLORYELAEAFNKTISTPKDODLIKQY--NTLIEBMLQVLIYVG 757
Db 797 LOCCALAPADSVYDKLSRFGLS-----SYLSLNPDTITNEYVPTVLLLEMLGLLIQ 851
QY 758 ERYVPGVGNVTKEEVTMRIIHLCLIEPMPHSAIAKNLPENENNETGLENVINKVATFKK 817

Db 852 ERRFCGLSTA-----ESLRREIIFKLATGDFTHSOLYKSLPRDLKSDELQEVLDVSVYCN 908
QY 818 PCVSHGVYELKDESLKDPNMTFYHYKQSKAEHMQKRRKQENKDEALPPPPPEFC 877
Db 909 PSGMNGKYSLQSSCKWELDLY-----HPRWQSRDLQSAEERF-----SRYC 950
QY 878 -----PAFSKVINLNCNDIMMYILRTVFEARAIDTDSNLWT-----EGMLQM 918
Db 951 GVSALTTLQPRWRMTYPLKGLARICTKATPQIISSALYALQSGTSVKSRAPOGLVIT 1010
QY 919 AFHILALGLEKQOLQKAPER---EVTDFYHKAS-----RLGSSAMNIQMLLEKLAG 969
Db 1011 ALQLLSLSDICTQORQNSQCCLENSIPLELAGLEIGTAQGTKEKSLLSLLVSLMK 1070
QY 970 IPQLEBQK-----DMITWT---LQMFDTV-----KRLREKSLIVATTSGSESIKND 1014
Db 1071 TRMGDGRHQFFPEPGSCNISSWIGNLLKKFSAIDSVCMNLLQSLAPEVVGQSGFDKVMSSG 1130
QY 1015 ITHDKEAERKKAAPAAHLRQKIMQAQSAQKQNIETHKLMYDNTSEMPGKE---DSI 1070
Db 1131 TSDEKRAKAKERQAA-----IMAKMAEQSKFLSTLSSMDD--DDPRSEFETSDSV 1181
QY 1071 MEEESTPAVSDYSRIALGP-KRGP-----SVTEKEVLTC 1103
Db 1182 MEHDEIAVREVCSLCHDPDSKDPVSFLIFLQVGMTDAIICDC 1225

RESULT 8
S64851
probable membrane protein YLR024c - yeast (*Saccharomyces cerevisiae*)
N:Alternate names: hypothetical protein LI730
C:Species: *Saccharomyces cerevisiae*
C>Date: 01-Aug-1995 #sequence_revision 24-May-1996 #text_change 19-Apr-2002
C:Accession: S64851
R:Obermaier, B.; Piravandi, E.; Rinke, M.
submitted to the Protein Sequence Database, May 1996
A:Reference number: S64845
A:Accession: S64851
A:Molecule type: DNA
A:Residues: 1-1872 <ORF>
A:Cross-references: EMBL:Z73196; NID:g1360331; PID:e245500; PID:g1360332; GSPDB:GN00012;
A:Experimental source: strain S288C
C:Genetics:
A:Gene: SGD:UBR2; MIPS:YLR024c
A:Cross-references: SGD:S0004014
A:Map position: 12R
C:Keywords: transmembrane protein
F:189-205/Domain: transmembrane #status predicted <TM1>
F:804-820/Domain: transmembrane #status predicted <TM2>
F:889-905/Domain: transmembrane #status predicted <TM3>

Query Match 5.9%; Score 535.5; DB 2; Length 1872;
Best Local Similarity 19.6%; Pred. No. 1.5e-24;
Matches 383; Conservative 320; Mismatches 721; Indels 529; Gaps 93;

QY 104 CGRVFKSGETPYSCRDCAIDPTCVLCMDQFQDSVHKHRY--KMTSTGGGCDGCDTEA 161
Db 98 CTRLCGFPSETIYCTCTNPLYEICELCFQKEKHVNHSYAKVVMRPEGRICHGCD--- 154
QY 162 WKTGPFVCHPGRAGTTK-----ENSRCLNCEVIVQARKIPSVIKYVEMPI 211
Db 155 ----PFAFN-DPSDAFKCKNELNNTPISDNSNVTTDDENVI-----SLLNYLDFLDIVTV 205
QY 212 -WEEKEL-----PPELQIREK-NERYKC-VLFNDEHHSYDHVIYS----- 249
Db 206 SYKEAEASHSERKASSLMHPNQNSITDDIMEKHECEPLVNDENVFEDNNKSNRKEAH 265
QY 250 LQALDCELAQLHTTAIDKEGRVRAKAGAACQAEKEDIKSH-----SENVQ-H 301
Db 266 MEWAQIEECNVHMOLDASTITRLNTPVEYAISITKALEDSDHVVTVLQSENFEED 325
QY 302 PLHVEVLHSEIMAH-----QKFAIRLGSWMNKK 329

Db 326 QIAKEFEQENIVVHRKADDIFKRKLTDDTLDMWLSLCFKAATSLQNKYALRI-SMLDVW 384
QY 330 MSYSSDFRQIFCOACUREPDSNP-----CLISR----- 359
Db 385 YSHFSK-----MRVSPNTNPDSKINLLGGFLISNEDSDSWFKPWSLENIEDER 435
QY 360 ----LMLWDAKL-----YKGAR-----KILHELIFSSF 383
Db 436 ISKILTYNBERLIRAHSPNTVSHFYNYGSRFQYIIINSINILSKSKFKML-KIMASLF 494
QY 384 FMEWEYKLFAMEFVYKQLOKEYISDDHRSISITALS--OMETVTPPLARHLIE-- 438
Db 495 SLRDESKFLAAQYIDVLSVLYDAVASD-AKEQVTLMSILGQYTFQDPSIANMTISSG 553
QY 439 --BQNVISVIT-----ETLEVLPEYLDNRNKNFNFOGYS--QDKLGRVAVIC--DLKY 486
Db 554 FIERTTRFAFTLMAFPNEDLMSYLPISL-----YNGFKLPTTIRNRTIICFKDLCT 606
QY 487 IL-----ISKPTIWTIRLMQFLEGFRSLKILTCMQGMEIRROVGOHIEVDP-DWEAAI 541
Db 607 INSANTVPELLSNEAIFNAIESFSESNVL-----PLKRETKHEVEVENFDFS--- 656
QY 542 AIQMQKLNILLMFOEW-----CACDEELL-----VAYKECHKAVM--RC 579
Db 657 AFYFFESSILIMTDGYTRISLVKDAFRKQIVLKLDDVAQTRFESLTNSRKKAIPDNA 716
QY 580 STSFISSSKTVVOS-----CGHSLETKSYRVEDLSVSIHLPLSLRTLAGHLVRLSLG--- 631
Db 717 STNENDSNKATLSTVRETICNYVAETINFGVNTQYFFNPMS-YLKFVYLOWQCGRYE 775
QY 632 -----AVSRLHEFVSFEDFOVEVLV-EYPLRCLVLVAQVVAEMRRNGLSISOVFYQD 685
Db 776 PIPASLTNYINLFEVQDKQKALYSISALSTLVLIQGINVGFWRNGTPTTHQARMYTK 835
QY 686 VKCREMYDKDIIMLOIGASLMDPNKPLLLVLQRYELAAEFNKTISTKDDOLIKOYNFLI 745
Db 836 YSMREFTYISDIFNVQFSMAKNPDELMTVYLSRWGLKHWAN-GVPMYDYPDPTTAVAV 894
QY 746 EEMLOVLIYIGE-RVYPGVGVNTKE--EVTMR-EIHLICIEPMPHSAIKNLPENENN 801
Db 895 NECILLIQLLTVEVRSI--VYMSKSGEFTFKSEIIHALCFDTCFAQIVNCIPEHITK 952
QY 802 ETGLENVINKVATFKKP-GVSGHGVYELKDESLKDFNMVYHYKTSQHSKAEH-----M 854
Db 953 HPSFDIYLEKANYTSPVSLTDNGIFVLKEKYKDEIDPYIIGLSSSRDYVEKNIRLMA 1012
QY 855 QKKRRKQENKDEALPPPPPEFCPA-----FSKVINLNCNDIMMYILRTVFEA 903
Db 1013 NLKMKMYED-----TFVPAKKVKOLLKNTLPSGLYSISSVNTFGLFLKNTLDHI 1061
QY 904 IDTDSNLTWTEGMLQMAFHILALGLEKQOLQKAPEEE---VTFDFYHKASRLGSSAMNI 960
Db 1062 IKYD-----YDNLPLRVVHLIHLVCMNVLNFMFGILMHEYAIVDTEFCYHYS-IGS--ILY 1114
QY 961 QMLEKLKGIPOLEGO-KDMITWILQMFDTVKRLREKSLIVATTSGSESI---KNDKI 1015
Db 1115 YCLLK--DNFSESHGKIREIFRYLM---ETAPHNVNVSYLREQTSYTPGILMWTKEKDS 1169
QY 1016 THDKEAERKKAAPAAHLRQKIMQAQSAQKQNIETHKLMYDNTSEMPGKEDSIMBEES 1075
Db 1170 HRDK---EPERKXHLARLKKLKKLAQOQMFENNSV---DTSOI-----S 1212
QY 1076 TPAVSDYSRIALGPKRGPVSVEKEVLT-----CILCQ--BEQEVKI-----ENNA 1118
Db 1213 TPRTTSPS---LSPTRINAENSSNTINSCDDDCVFCCKMPKDDDFVYFVSYOERNICDHG 1269
QY 1119 -----MWLSACVQKSTALTOHGRKPIELSGEALDPLFMDPDLAYGTYTSCG 1165
Db 1270 IDFTNPTDVRNIRNSLFSFGKOTKDSAOEN---PQDDGDTRLKFTSCPEVLR-----ACG 1320
QY 1166 HVMHVAWCQKYFEAVO-LSSQQRTHVDLFDLESEYILCPCKSKSLCNTVIP----- 1214
Db 1321 HGSHTKCLSGHMKSIKGIQNTKNTPL-SYSGSLIYPCVNSLSNSFLPKTNDIDKRTS 1379

Qy	1315	---IIPLOQKINSNADALAOQLTLARITQTVLARISGVYNIHAKGNPIPIFNQMG	1271
Db	1380	SOFFMCIEKRSEABENLDPMSGICAKAMI-----LG	1411
Qy	1272	DSTLEFHLSILSGVESSIKYSKEWVILFATYIRIGLKVPPDERDPRVPM	1329
Db	1412	D--LOGKKVTT--IDAYKVVNSV-----FINTSNTELRLRSHKKEGKIVNMERISQ	1461
Qy	1330	CAFTTQAIENLIGDEGKPLFGALQNRHNGLKALMQFAVAQRI--TCPOVLQIKHLVRLLS	1388
Db	1462	CILTLHLVCELKSFYIKKRVNS-----KTFSEISRKIWNNEFLIKGNVNLL	1511
Qy	1389	VLPNI-----KSEDTPCLLSIDLFH-----VLGVAVLAFPSLYWDDPVDLPQSSVS	1435
Db	1512	YMSQNFNDIDGKTPOPPNLCIYEMKRRFHQLLLLLLARDMVRVNFYKDCRNKIKISSNG	1571
Qy	1436	S---SYNHLHYLPHLITMAHMLQILLTVDTGCLPLAQVQDESEEAHSASFFAEISQYTS	1492
Db	1572	SEEPSTSFYLFN--TFKRYVDFRPDD-----VRFDFTSLEKIKDFICSL-----	1615
Qy	1493	IGCDIPGWVWVSLKNGITPYLRCAALFHYLLGVTPPEELHTNSA-----EGEYSAL	1545
Db	1616	-----LLESLSIFGRTEFLFNQYDDGDDGNNNRNSNFMFDMVKQREIELI	1662
Qy	1546	CSYLSPTNLFLLFOEYNDVTRPQLQ-----RCADPALLNCLKOKNTVVRPKRN-	1597
Db	1663	FRYFKLPNLTHFLKOFFYNELQNIERYNGDNDNLRIQQVIYDMVQNTINTRA-VPSPEH	1721
Qy	1598	SLIELPDDYS--CLLNQASHFRCPRSADDERKHIPVLCILFCGAILCS-----QNI	1644
Db	1722	QLIELPLNLSKFSLDNDEISNKC-----DKYETIAVCLLCGQ--KCHIQKSIALOGYLO	1774
Qy	1645	CCQETVNGEEVGAC--IF-----HALHCK--ARGCAYPAPVLDYGETDPLKRGKNPLHL	1695
Db	1775	CTDHRNCEIITSAYGVFLMTGNALISYVKKRGTFYAAPLYLSKYGETNEDYKFGTPVYL	1834
Qy	1696	SREYRKHLVWQOHCIIIEIARSQETNOMLFG	1728
Db	1835	NRARYANL-----ANEIVFG	1849

RESULT 9

D88131

C:Species: Caenorhabditis elegans

C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001

C:Accession: D88131

R:anonymous, The C. elegans Sequencing Consortium.

Science 282, 2012-2018, 1998

A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology

A:Reference number: A75000; PMID:99069613; PMID:9851916

A:Note: see websites genome.wustli.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C_elegans/

A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and

A:Accession: D88131

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-2168 <STO>

A:Cross-references: GB:chr_II; PIDN:AA81132.1; PID:g1055163; GSPDB:GN00020; CESP:F10G7.1

C:Genetics:

A:Gene: F10G7.10

A:Map position: 2

Query Match	5.6%	Score 507.5;	DB 2;	Length 2168;
Best Local Similarity	18.2%	Pred. No. 1e-22;		
Matches 393;	Conservative 333;	Mismatches 772;	Indels 663;	Gaps 87;

Qy 18 EISAELPQTPQRLASWMDQVDYFATF--LHHLAOLVPEIYFAEMDPDLKEQESVOMS 74

Db 214 DLAELEKSNYPYA---EGIDNYQSFDETRKKAHMLDDFLDGFIDELEQDEKTFEIL 269

Qy 75 IFTPLLEWYLFGEPPDICLEKHLKSHGAFQGLCGRYVKSGETTYSRCRDCAIDPTCVLCMDCQF 134

Db	270	RVLIAQ----	-GEPYMEFKQKMKENDPSLKCNETWENDAVAFRCNTCALTFCMSLCDCEPCE	325
Qy	135	DSVHNHRYKMHSTGGGFCDCGDT	PAWKTPGCVNHNHPEGRAGTIKENSRCPLNEEVI	194
Db	326	SNHGAGHDYTRFTRREGGACDCG	NODVIKBOGNCKNHG-----	378
Qy	195	ARKIPPSVIKYVEMTITWEEKE	LPELQITREKNERYCYVLFNDEHSHSDYHDVYISLQ	254
Db	379	AEYI--VMKLIVRLF-----	LEYRGWASRY-----	405
Qy	255	DCELAEAQLHTTADKEGRAV-	KAGAYAACQAKEDIKSHSENVSOHPLHVEVLHSEIM	313
Db	406	D-----EFSESNEP	INRRAVDIGFTVDDAARSDY-----	437
Qy	314	AHOKFALRLGSMWNKIMSYSD	RFQIPQACLAERE-----	365
Db	438	-----LITFLOECVKY	GGPMRLIVSKILLDKNLKYKVLTEQTTNRRVOIS--	487
Qy	366	KLYKGARK-----	ILHELIF-----	387
Db	488	HMFKDDRTVLPPD	NSFNFSVGSTESTECTSLDDELIFWINRLOPLQNLVNFGLSLLSEP	547
Qy	388	EYKLFAMEFVKYKOLKE	YISD-----	433
Db	548	GYRDAFAIRFESWY-PISG	KVISDLCISQAGHGRADOLVTPACSRVHVTVQMLSGASLC	606
Qy	434	RHLTEQONVI-----	SVITETLLEVLPEYLDNRNKNFQGYSDQKLGVR--	477
Db	607	KELNENVHLKTI	FDVTRFLFICEKSVDSLSKSNIFKRESSRFLNMTTNGRPKRWMT	666
Qy	478	-----YAVICOLKY	ILISKPTIWTBRLMOF-LEGF-RSFLKILTCOMQMEEI	523
Db	667	MANNAISQHGVPFVMDIQ	NVL-----	722
Qy	524	RRQVQHLEVDPDWEAAIA	IQOMQLNLLMFQEWCAACDEBELLVAVKECHKAVMRCSTSP	583
Db	723	WRIISGNAQENDAGEA-----	QRAYTLEFETLAVTFNIVAA-----	760
Qy	584	ISSKTVVQSCGHSLET	KSYRSVEDLSIHPLSRT-----	637
Db	761	IQEKKVRTAC-----	RFDHIRSOHISITAITHTFELDCGYDHIRSNLLN-----	805
Qy	638	EFSFEDQFQVEVLVEYPL	CLVLVAQVVAEMWRNGLSLISQVYFYQDVKCREMYDKDI	697
Db	806	-----DETLRLLLVHPL	RIQVAAEINCNNWYRNGAQAAMKSAIYSOMNVSSAFQTPDV	860
Qy	698	IMLOIGASLMDPNK	FLLLVLRQVELARAF-----	726
Db	861	DLIRFCAAHIDKEHFM	AAHKTASNITECIKIQGRFVFEKSESKILFQGVGGDDDEVSE	920
Qy	727	NKTISTKQODLIKQYNTL	IEEML-----	754
Db	921	ARKTEFERYQLVRE	QALIMEDMMLTPKLNLFQFDEKETDEADVMIENNATEVITDFT	980
Qy	755	-----IVGE-----	RYVPGVG-----	765
Db	981	IQSPIDHSRDPRI	PTVGEFIRRHLEAAGVADIDVEMEREFDRFLFDDEREVTIVIREP	1040
Qy	766	-----	NVTKEVTHREIHLICIEPMPHSAKTNLPE-----	797
Db	1041	EWIDPMFWMGPKLVAEL	VVVRVNSGASPEEHYKSEMYNCNAMGNVYSLRSLASISKEGSR	1100
Qy	798	-NENNETGLENV	INKVATFKKP-GVSGH-----	850
Db	1101	GSEMDKHFEFILNE	IGDFTPIETTHLMOGSGYQLK-TSINSEVCPVFFMRSRSIKQ	1159
Qy	851	AEHMOKKR--RKQEN--	KDEALPPPPPPFCP-----	892
Db	1160	AREVFAKMEIREQ	NALDKOLEVAKIEERFWPFLRDFSDQKRRHGISNLYNVLLTERF	1219
Qy	893	MYILRTVFERAID	TSDNLWTEGMLQMAFHILAGLLEEKOOLAKAPBEEVTFFDYHKASR	952
Db	1220	LIHCITVILASPAD-ETAK	FHDGTQVLAVLITLGVKYAQSVGDEKIKKOMIDIFTTFQ	1278

QY	953	-----LGSSAMNIOMLLEK-----LKIPOLLEGOKDMIT-----WILOMF	987
Db	1279	LIKPREMETFTVCAFMIRLLTKETRKGALVVVFKGILSGEYDKERVTKGKIYLARFV	1338
QY	988	DTVKKRLKSCSLIVATTSGGESIKNDIETHDKAE-----RKRKAEEARLHROKIMAQ	1041
Db	1339	TILTKLSPVAROIIIEGKLKEELRISKHSRNOEKMKAPMDPVKAAKAAKARMEALM-----	1396
QY	1042	MSALQKNFIETHKLMYDN-----TSEMPGKEDSIMBEESTPAYSDYSRIALGPKRGPSPVTEK-	1098
Db	1397	QNSAKKSAQTMELMKMTEGTDAEVKNKVDPSQNRKV-----YECPICGEONAPVTENP	1451
QY	1099	-----EVLTCILOEBEVEVIENNAVL-----SACVOKSTALTQHRGKPIELSGEA	1145
Db	1452	FGMLAKLSTNICEBOIDASINTIDDLKFEDEYHVSANLQSETRRRFFSKRQATFEN	1511
QY	1146	LDPLFMDPDLAYGTYTSCGHVHVAWCQKY-----FEAVQLSSOORIHVDLFD-----	1194
Db	1512	QDIVKVNPLV-GDGLTKCGHAAHACFNAYRASLVSGFGEKKLENEKK-----KYDGHQR	1566
QY	1195	LESEGYLCPLCKSLCNTVPI-----IPLQPKINSENADALAQLTLTLARWTQTVLAR	1247
Db	1567	STDRREVGCCMRYTVNVIVPMSEDKPYTIKTPASPMYSDV-----W-----R	1611
QY	1248	ISGVNIRHAKENPIPIFFNOGMDSTLEFHSILSGVESSIKNY--STIKEMVILFATT	1305
Db	1612	VMDVVLKKKAG---PVFOED-----ERNVKYSTNYSTREGGLFELY	1650
QY	1306	IYRI-----GLKVPDDEPRVPM-----L	1325
Db	1651	IGRHSADLAERKSSQOQCTVSTLWVSLAVVMVERSSIMRKMGPYRKRKNORMSIMEHLM	1710
QY	1326	TWSTCA-----FTIOAIENLLGDECKPLFGALQNRHGNLKAALMFAVAQRITCPOVL	1378
Db	1711	TASVATSKDVFDFVALSALTNLFAKVTESESRPPSEQPSTKEPAQGTAPN-----PEAV	1767
QY	1379	IQKHLVRLLSVPLNKSEDTPCLLSIDLFEV-----LVGAVLAFPPSLWDPDVLDQ	1430
Db	1768	VGLSSDEMAAMITKPLRKDTPDCLSKLPFLALDPKATLVRLMAVLIDNQSLTKDIOREIA	1827
QY	1431	PSVSSSYNHLIYFLHLITMAHMLQILLT-----VDTGLPLAQ-----VQEDSEEAHSA	1478
Db	1828	QNMIFASLGNVWSYTLCL-----ILRTGEKISALNKGEPKIOGLSDHILQSEAEACQA	1882
QY	1479	-----SSFPAEISQVTSIGCDIPGWLWVSLKNGITPYLRCA-ALFFHYLLGVTP-PEE	1532
Db	1883	LTYNTEFKHLAQRLE-SPDVEPTDEYICKTMSCLIEFLRFSEVLLFHCNLGENDVNNQ	1941
QY	1533	LHTNSAEGE-----YSALCSYLSLPTNLFLLFOEYND-TYRPLLQ-----RRCADPALLN	1581
Db	1942	IHSRNIDLNVRLRVGINAQNINVPKG-----ANYWTFRFTQLTNMAKLRIFQPCIVE	1996
QY	1582	CLKOKNTVRYPRKRNLSIELPDDYSCLLNQAASHFRCPRSADDERKHVPVLCILFCGAILCS	1641
Db	1997	PLAWK-----PRR--ILKPPTNFDELFGRYFHREC-NKCSSVPSPVPMCLFCGEILC-	2045
QY	1642	QNTICQ-----EIVNGEEVCACIFHALHCKA-----RGCAYP	1673
Db	2046	LNDCCRMAHOGESGDRVISMSEIEA---HAEDCSSSGGLFSLVTSSMVVYSRGKQAAIWG	2102
QY	1674	APYLDGEYDTPGLKRNPLHLRSERYRKHVLVWQHCIIETIARSOE-----TNQMLFGF	1729
Db	2103	TIYLDHAIKEEDRNLRKCKPLFELCESRLKWLKEYDWAEE---QEWQRVPQWFNTNSQAFTS	2158
QY	1730	N	1730
Db	2159	N	2159

RESULT 10

T48252

eceriiferum3 (CER3) - Arabidopsis thaliana

N: Alternate names: protein T1E22.70
C: Species: *Arabidopsis thaliana* (mouse-ear cress)
C: Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C: Accession: T48252
R: Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft
submitted to the Protein Sequence Database, March 2000
A: Reference number: 224489
A: Accession: T48252
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-795 <BEV>
A: Cross-references: EMBL:AL162874
A: Experimental source: cultivar Columbia; BAC clone T1E22
C: Genetics:
A: Map position: 5
A: Introns: 203/2; 220/3; 428/1; 488/3; 666/2; 701/2; 723/3; 749/3; 772/3
A: Note: T1E22.70

Query Match	3.28;	Score 292;	DB 2;	Length 795;
Best Local Similarity	21.6%;	Pred. No. 5.2e-10;		
Matches 180;	Conservative 117;	Mismatches 265;	Indels 270;	Gaps 42;
<hr/>				
Qy	1038	IMAOMS----	ALOKNFETIKLMY-----	DNTSEMPGKEDSTIMEESTPAYSDYSRIA 1086
		: : :	: :: :	: : : :
Db	56	LMLQLSDDTISESANMIESTKARLIGNQTEKRSGDGKDSEWNE-----	SLEIATMYOTVR 112	
Qy	1087	LGPXRGFSVTEKEVTLCILCOEEQEVIKENNAMYLSACVOKSTALTQHRKPIELSG--	-- 1143	
		: : : : :	: :: :	: : : :
Db	113	-----NKINEMINQSILTRVDHQHAEAN-----	CSEKNSV-----	GGPSTLOGREP 153
Qy	1144	-----EALDPLFMDPDLAYGYTGSCGHVMHAVCWQKYFAEVQLSSOORI 1188		
Db	154	DIRSRQTSRRPDAGSDGFIDCO-----	GVYLSSCGHAVHQSLERLYLSLKERSGRGT 208	
Qy	1189	HVD---LFDLSEGYELCPCKSLCNTVIPPI-----	LOPO-KINSENADALAOLLT 1236	
		: : : : : :	: :: :	: :: :
Db	209	VFGAHIVDLKKREFLCPVCRRLANSVLPECPGDCLSVKLQDSPRTKLRRKDALQPSL-	- 267	
Qy	1237	LARIWTQTLARISGYNIRHA-----	KGENPIPIFFNOGMGSTLIEPHSIILSFVGVESS 1288	
		: :: :	: : :	: :: :
Db	268	---WLSEALCL-----	LRSAAEVIEDGRKTVTP-----	QGQGPRR----- 301
Qy	1289	IKYSNSTIKEMWILP-----	ATTIYRIGLKVPDPDERPRVPMLTWSTCAETIOAIENLIG 1342	
		: : : : : :	: :: :	: :: :
Db	302	-KDKUSVKMLWDYFPKPKDETKURLWL-----	PPOS-----	IYMWDTIKYLSLISME-- IG 350
Qy	1343	DEG-----	KPILEGALQNQRHNGLKALMOFAVAOR---	ITCQPVLII-QKHILVRL 1386
		: : : : : :	: :: :	: :: :
Db	351	TRPAKNMPLPYVICDSLSEELKTSGTILSVLLRVQSSRTKNTIHVRQFVGMKHLAES	410	
Qy	1387	LSVVLPNIKSED-----	TPCLLSIDL-----	FHVILGAV--LAFPSLYW 1423
		: : : : : :	: :: :	: :: :
Db	411	ICYGVSSSSSSIFSGEGTTGSLKNIDILLNNRASDPVLAHDPFSSLMAWLFCLUPFPFLTC	470,	
Qy	1424	DDPVLDQSPSSVSSNYHLFIHLITMAHMLQILLTVDGLPQAQVQDSBEAHSASSFFA	1483	
		: : : : : :	: :: :	: :: :
Db	471	EESL-----	LSLVHFHSVSLVQTVIAYCACRP-----	SELSELNFGENLLN 512
Qy	1484	EISOYTSGS-----	ICGDIPGWYIMWSLKNGITPYLRCAALFYHLLGVTPPE 1531	
		: :: :	: : :	: : : :
Db	513	DISNALRESGGWEYFRSNNDLSCDIKD----	TIRKYSLPFLRRCALLMKLL--	KSTPR 565
Qy	1532	ELHTNSAGEYSALCSYLSLPT--	NLEFLFEQ-----	EYWDIVRPLLQRCADPA 1578
		: :: :	: : :	: : : :
Db	566	KLH-----	BESDMFDLPSPDTTDNMDFITYSPQSELNHVQELKMFNIPPI-----	DII 613
Qy	1579	LLNCLKOKTNW-----	RYPKRNS-----	LIELPDDYSCLLNQASHFRCP 1619
		: : : :	: :: :	: :: :
Db	614	LNDELLRSSTQIWLOHPQREYRNVRNRKSICITPVVPFOLMKPLPNLYODLLQRCIKRCV	673	
Qy	1620	RSADDERKHPVLCFCGAILCSQ--	NITCCOEIVNGEEVGACIFHALHCK-----	1666
		: :: :	: :: :	: :: :
Db	674	-NCTKVIEEPVCLLCSG-LCSPILWSPCCRE-----	SGCPNHAIPTCGAGTGVFLLIRR 724	


```
Qy 1667 -----ARCAVPAPVLDYGETDPLKRGNPLHLRSRERYKRLHLVWQOH 1710
Db 725 TTILLQRFAROSPWPSPYLDFTGBEDIDMRKRLYLNEERYAALTLYLVGSH 776

RESULT 11
T40238
hypoetical protein SPBC32F12.14 - fission yeast (Schizosaccharomyces pombe) (fragment)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T40238
R:Moreno, S.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, June 1998
A:Reference number: Z21915
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Accession: T40238
A:Molecule type: DNA
A:Residues: 1-271 <MOR>
A:Cross-references: EMBL:AL023796; PIDN:CAA19375.1; GSPDB:GN000067; SPDB:SPBC32F12.14
A:Genetic source: strain 972h; cosmid c32F12
C:Genetics:
A:Gene: SPDB:SPBC32F12.14
A:Map position: 2

Query Match 2.4%; Score 218; DB 2; Length 271;
Best Local Similarity 26.0%; Pred. No. 4.le-06;
Matches 60; Conservative 34; Mismatches 97; Indels 40; Gaps 6;

Qy 65 EKQESVQMSITPLEWY-----LFGEDPDICL-----EKLKHSGAFQIC 104
Db 37 ESAKSLNLFVFSALLGYDHTLWTLTPERTIDASFLLRRAQHGSEGEVRRHGTCSEKC 96

Qy 105 GRVFKSGETYSRCDCAIDPTCVLCMDQFDSVHKHNRHYKMTSTG-GGFCDCGDTAEWK 163
Db 97 GHIFRKGVEFYRCKTSVDSNSALCVKCFRATSHKHDETSFVSAGSGCCDCGNAAWI 156

Qy 164 TGPCVNHPEPRAGTIENSRCPNEEVIVQARIFPSVIKYVEMTWEET--KELPPE 221
Db 157 GDVSKIHSHEDATISNDMIDEIPEKLENSIQTTIDCVLDVLDVDFVSCSPENLKKMPTL 216

Qy 222 LQI--REKNER-----YYCVLFNDEHSHYDHYVYLSQRLAD 255
Db 217 ESILQDEKTSRLSENKYGDIDDSNMYSLVLWNDEKHSFKFYEQYITTALE 267

RESULT 12
JC5837
364K Golgi complex-associated protein - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 05-Mar-1998 #sequence_revision 13-Mar-1998 #text_change 20-Jun-2000
C:Accession: JC5837
R:Toki, C.; Fujiwara, T.; Sohda, M.; Hong, H.S.; Misumi, Y.; Ikehara, Y.
Cell Struct. Funct. 22, 565-577, 1997
A:Title: Identification and characterization of rat 364-kDa Golgi-associated protein rec
A:Reference number: JC5837; MUID:98093490; PMID:9431462
A:Accession: JC5837
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-3187 <TOX>
A:Cross-references: DBJ:D25543; NID:g516825; PIDN:BAA05026.1; PID:g516826
C:Comment: This protein plays a role in the formation and maintenance of the characteris
C:Superfamily: giantin
F:49-549,624-1176,1238-1707,1763-3114/Domain: coiled-coil leucine zipper #status predict
F:3165-3187/Domain: membrane anchor #status predicted <MAD>

Query Match 2.2%; Score 205.5; DB 2; Length 3187;
Best Local Similarity 20.2%; Pred. No. 0.00098;
Matches 243; Conservative 182; Mismatches 472; Indels 305; Gaps 49;

Qy 130 MDCFQDSVHKHNRHYKMTSTGGGFCDCGDTAEWKTPFCVNHPEPRAGTIK-----ENS 183
Db 1057 IDLLQEEITENQATQIKFITGTMDAGDGDGSAVKETSV-----SSPPRAGGGEHWKPELECK 1112
```

```
QY 1060 TSEPGKEDSIMEESTPAVSYSRIALGPKR-----GPSVTEKEVLTCLCOEQEVKI 1114
Db 2127 TKSMSSLDQ-----DRDRVIDEAKKWEQRFDAIQTK-----EEVRLKE 2166
QY 1115 EN 1116
Db 2167 EN 2168

RESULT 13
A56539
N:Alternates: human
C:Species: Homo sapiens (man)
C:Date: 19-Oct-1995 #sequence_revision 26-Jan-1996 #text_change 10-Dec-1999
C:Accession: A56539; S37536
R:Seelig, H.P.; Schranz, P.; Schroeter, H.; Wiemann, C.; Griffiths, G.; Renz, M.
Mol. Cell. Biol. 14, 2564-2576, 1994
A:Title: Molecular genetic analyses of a 376-kilodalton Golgi complex membrane protein
A:Reference number: A56539; MUID:94187728; PMID:7511208
A:Accession: A56539
A:Molecule type: mRNA
A:Residues: 1-3259 <SEE>
A:Cross-references: EMBL:X75304; NID:94057114; PIDN:CAA53052.1; PID:94057115
C:Genetics:
A:Gene: GDB:COLGB1; GCP: GCP371
A:Cross-references: GDB:454958
A:Map position: 3q13.31-3q13.31
C:Superfamily: giantin
C:Keywords: coiled coil; Golgi apparatus; transmembrane protein
F:3238-3254/Domain: transmembrane #status predicted <TM>

Query Match 2.1%; Score 190.5; DB 1; Length 3259;
Best Local Similarity 18.0%; Pred. No. 0.0086;
Matches 221; Conservative 207; Mismatches 424; Indels 377; Gaps 49;

QY 169 VNHEPGRAGTICKENSRCPNNEVTVQARKIPSVIKYVYVMTWIEEEKLPPELQ--IRE 226
Db 12 VLHLGGDDDDTDDNNRAPLDPHELQESDMFEFNTTQEDYQERLAYAE-QLVVVELKDIIRQ 70
QY 227 KNERYCYVLFNDEHSHSVHVIYSQALDCELAELQHT---TADK--EGGRAVKAGA 280
Db 71 KDVL-----QKDEALQEEKADNKKLKLHAKALTSUNKYIEEMKAGGVV 121
QY 281 YAAOQAEKEDIKSHSEYNSQHLHVEVLHSEIMAHQKFAIRLGSWMNKIMSYSSDFRQIF 340
Db 122 LPTEPQSEQLSKHDKSSTEREMEIEKIKHKLQEEELISTL----- 163
QY 341 COACLREPPDSNPCLISRLMLWDALKYKGARKILHELIFSSFFMEMEYKFLFAMEFVKY 400
Db 164 -QAQL-TOAQAEPQAQSS-----TEME-----EFVMM 188
QY 401 YKOLQKEYISDDHDSRISITALSVMFTVPTLARHLIEQN-----VLSVIT 447
Db 189 KQQLQEK-----EEFISTLQALQSQTQAQAAQVVRKEDARFETQVRLHEDELLQVLT 242
QY 448 ETLLEVLPEYLDNRNKNFNGYSQDKLGRVYAVICDLKILYS---KPTIWERLRLM-- 501
Db 243 QADVETEMQOKLRVLQRKLEHEESLVGR--AQVVDLLOQELTAEQRNQLISQQLQOME 300
QY 502 -----QFLEGRSFLKILTCMQGMEIEIRVQGHIEYDPDWEAAIAIQMKNLILMF 554
Db 301 AEHNTLRNVTETREESKILLEKMELEVAERKLSFH-----NLQEEHHLLEQF 349
QY 555 QEWACADEL--LIVAYECHKAVNRCSTSF--SSSKT--VVQSCGHSLETYSRYRVEDL 609
Db 350 EQAGQAQAELESRYSALEQKHKAEMEKTSHLSLQKTQELQSACDALKDQNSKLLQDK 409
QY 610 VSIHPLSRTLAGLHVRLSLRGA-----VSRL-----HEF---VSPEDFOVEVLVEYPLRC 657
Db 410 NEQAVQSAQTIOQLEDQLOQSKSEISQFLNRLPLQOQHEFASQSPFDVYNQGTAVTEEN 469
QY 658 LVLVAQVVAEMWRNRGLSLISQVFFYQDVYKREMYDKDIIMLQI-----GAS 705
```

```
Db 470 IASLQKRVVELENEKGALLSSI--ELEELKAENKLSQITTLLEAQNRTGADREVSVIS 528
QY 706 LMD-PNK-----FULLVLQRYELAE 725
Db 529 IVDIANKRSSSAEESGQDVLENTFSQHKKELSVLLEMKAEQAEATAFLKLQLOGRABEA 588
QY 726 FNKTISTKQDQLIKQY--NTLIBEMLOVLIYICERY----- 760
Db 589 DHEVL---DQEMKQMEGEGIAPIKMKVFLDGTQDFPLMPNESSLPAVEKEQASTEHQ 645
QY 761 -----VPGVGNVTKEEVTMTREIITHLLCIEPMPHSA- 790
Db 646 SRTSEELISNDAGVELKSTKQDGKSLSAVPDICOCHQDELE-RLKSOILELELNFHKAQ 704
QY 791 --IAKNPENNETGLENVINKVATFKPGVSGHGYV---ELKDESL---KDFNMYF- 840
Db 705 EIVEKNLDEKAKEISNLQLEE--PKKNADNNSAFTALSEERDQLLSQVKELSMYTE 761
QY 841 --YHYSKTQHSKAEHMQKRRKQBNK---DEALPPPPPPPCFPAPSKVINLLNCDIMMYI 895
Db 762 LRAQVKQLEMLNLAEAERQRLDYESQTAHDNLLTE-----QIHSLSIEAKSKDVKTEV 814
QY 896 LRTVFE-----RAITDSNLWTEGMLQMAFH 921
Db 815 LQNELDDVQLQFSEQSTLIRSLQSLQNKSEVLEGAERVRHHSKVEELSQALSQKELE 874
QY 922 ILALG--LLEK---OOLQKAPER-----EVTDFYHKASRLG---SSAMTQMLLEK 966
Db 875 ITKMDQLLLEKKRDVETLQOTIEEKDOQVTEISFMTKMKVQNLNEEFSLQVEIKTLKEQ 934
QY 967 LKGI-----POLEGQKDMITWILQMFDTVKRLREKSLIVATTSGESIKNDEITHD 1018
Db 935 LNLLSRAEEAKKEQVEDNEVSSGLKQNYDEM-----SPAGOISKEELQHE 980
QY 1019 -----REKAPRKKAEEAARLHRQKIMAOMSALQKNFIETHKLMVDNTSEMFGKEDSIMEE 1073
Db 981 FDLKKENQKRRKLAALINRKELLQVRSLER---ELANLKDESKKEIPLSETERGEV 1037
QY 1074 ESTPAVSDYSRIALGPKRGPSTVEKEVLTCLCOE-----EDEVKTIENNAWLVS- 1122
Db 1038 EEDKENKEYSE-----KCVTSK-----CQEIYILYKQTISEKEVELOHTRKDLEE 1082
QY 1123 --ACVQKSTALTQHRGKPIELSGEALDPL 1149
Db 1083 KLAAEQFQALVKOMNQTLDKTNQIDLL 1111

RESULT 14
A57013
N:Alternates: antigen 1 - human
C:Species: Homo sapiens (man)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A57013; S44243
R:Mu, F.T.; Callaghan, J.M.; Steele-Mortimer, O.; Stenmark, H.; Parton, R.G.; Campbell, J. Biol. Chem. 270, 13503-13511, 1995
A:Title: EEAL, an early endosome-associated protein. EEAL is a conserved alpha-helical
A:Reference number: A57013; MUID:95286647; PMID:7768953
A:Accession: A57013
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1410 <RES>
A:Cross-references: GB:L40157; NID:g1016367; PIDN:AAA79121.1; PID:g1016368
R:Seelig, H.P.
submitted to the EMBL Data Library, April 1994
A:Reference number: S44243
A:Accession: S44243
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-254,'C',256-257,'LQ',260-276,'A',278-283,'A',285-519,'D',521-574,'EQ',5
A:Cross-references: EMBL:X78998; NID:g475933; PIDN:CAA55632.1; PID:g475934
C:Genetics:
```

A;Gene: GDB:EBA1

A;Cross-references: GDB:l1369996

C;Superfamily: human early endosome antigen 1

C;keywords: calmodulin binding; endocytosis; metal binding; peripheral membrane protein

Query Match 2.0%; Score 186; DB l; Length 1410;
Best Local Similarity 18.5%; Pred.No. 0.0047;
Matches 204; Conservative 200; Mismatches 424; Indels 272; Gaps 45;

QY 181 ENSRCPUNEVIIOARKIPFSVIVKVVEMTWEEKEKLPPPELOIREKNERYVCVLFNDEH 240
| : | : | : | : | : | : | : | : | : | : | : | : | :
Db 275 ELSKGQEVRVYVEQLOKLKSSVNELTKNQTLTENLLKKEDYTCLKLEKH-----NEES 329
| : | : | : | : | : | : | : | : | : | : | : | : | :
QY 241 HSYDHVIYSL-QRALDCELAQAOLHTTAIDKEGRRAVKAGAYAAQCPEAKEDIKSHSENV 299
| : ||||| : | : | : | : | : | : | : | : | : | : | : | : | :
Db 330 VSKNKIQATLHQDXDLDCQQLOSRL-----SASE-TS 359
| : | : | : | : | : | : | : | : | : | : | : | : | :
QY 300 OHPHLHEVLHSEIMAHQKFALRGSMWNKMINSYSDPROIFCQACLREEPDPSENPCILSR 359
| : ||||| : | : | : | : | : | : | : | : | : | : | : | : | :
Db 360 LHRHV-B-LSKGEATGKLEEISEVTKYQHUKAEFKVQLQQO---REEKEHGGLQOSE 415
| : | : | : | : | : | : | : | : | : | : | : | : | :
QY 360 LMLWDARLYGARGKILHELIFSFFMEMEYKFLFAMEFYKVYKQLKEYISDDHDRSISI 419
| : | : | : | : | : | : | : | : | : | : | : | : | :
Db 416 INQLHSKI-----LETQRQCGAHLRKLRQSSEKLMWDRQQ--V 455
| : | : | : | : | : | : | : | : | : | : | : | : | :
QY 420 TALSVMQTPTTARHLIEQNVISVITEPLLVPYLORNKNFN-FGYSDCKLGIRV 477
| : | : | : ||||| : | : | : | : | : | : | : | : | : | : | : | : | :
Db 456 ADIQLK-----LSR-L-EBQLKEKVTNSTEQ---HOLDKTQOHQEOAQALOOSTTAKL 504
| : | : | : | : | : | : | : | : | : | : | : | : | :
QY 478 YAVICDLULYILISKPTITWRERMOLEGFRSLIKLTCHOGMEIERROGHLEVDPDW 537
| : ||||| : | : | : | : | : | : | : | : | : | : | : | : | :
Db 505 REANDUEQL---RIGEKDKXIONREAL-LQKSENISLLEKEREDYAKTIQG-EG 558
| : | : | : | : | : | : | : | : | : | : | : | : | :
QY 538 EAATAIQMQLNILLMPQEWACADEELLV-----AYKECH-----KAVMRC 579
| : | : | : | : | : | : | : | : | : | : | : | : | :
Db 559 ETAVLNQLQEKNTLQ-----DEVTLTENVKNOESHKQAGENLDVDQOEKAHLRA 611
| : | : | : | : | : | : | : | : | : | : | : | : | :
QY 580 STPFISSKTVWGSGHSLETSKSVRSDELVS I-----HLPL 616
| : | : | : | : | : | : | : | : | : | : | : | : | :
Db 612 AODRVLSLET SVNELSNQLNESKVESQDSIDIQIAKAKTELLISAFAAKTAORADLONHLT 671
| : | : | : | : | : | : | : | : | : | : | : | : | :
QY 617 SR-TLAGLVHRLSLGA-----VSRLHE----FVSFE-----DFQ 646
| : | : | : | : | : | : | : | : | : | : | : | : | :
Db 672 AQNALQDKHQELNAKITTDQDVYAKLDQEHCSQLESJHKYEKYLEUQETEELEGQ 731
| : | : | : | : | : | : | : | : | : | : | : | : | :
QY 647 VEVLVEYPLRCLVLVAOVAMRRNGLSITSQVFYYQDVYKCREMYDKDIIMLQ--IGA 704
| : | : | : | : | : | : | : | : | : | : | : | : | :
Db 732 IKKLEADSLEYKASKEQALDQOQROLN-----TDLELATELS KOLEMEKEIVSS 783
| : | : | : | : | : | : | : | : | : | : | : | : | :
QY 705 SLMDPNKFLVLIORYELAFAFNKISTKDQ--LIKQ-YNTLIEPMLVOLVIYVGRIYV 761
| : | : | : | : | : | : | : | : | : | : | : | : | :
Db 784 TRLDLQK-----KSPALESIIKOJLKBEEKOIJKODFTLSQE-TKIQHESLNNRIQ 835
| : | : | : | : | : | : | : | : | : | : | : | : | :
QY 762 PGVGNVTKEEVTMRITTHLICIEPMPSHSA TAKNLNPENNETHGLENIYNKVA----- 813
| : | : | : | : | : | : | : | : | : | : | : | : | :
Db 836 TTVELQGVNMKEPALMTSTVYKDKLSKVSDSL-KNSKSEFEKENGGKAAILDLEKTC 894
| : | : | : | : | : | : | : | : | : | : | : | : | :
QY 814 -----TFKKPGVSGHYVELKDES--LK-DFNMFYFHYSKTOHASKHAHMOKK 857
| : | : | : | : | : | : | : | : | : | : | : | : | :
Db 895 KELKHQLQVQMENTLKEQKELKLSLEKEKEASHQKLELNSMQBLQIAONTLQONEKEE 954
| : | : | : | : | : | : | : | : | : | : | : | : | :
QY 858 RRQENKDEALPPPFPFCAPSKVINLCNDIMMUIL-RVFERAITDDS NLWTGMGL 916
| : | : | : | : | : | : | : | : | : | : | : | : | :
Db 955 QOLOGMTNEL-----KOSSEQBKQTEALQUGELKIAVLQKTELLENKLQOO---LTQAAA 1005
| : | : | : | : | : | : | : | : | : | : | : | : | :
QY 917 OMAPHIALGLBLEKQOLQKAPEEVFDVFKASRLGSSAMNTOMLEKLUKPQILEGO 976
| : | : | : | : | : | : | : | : | : | : | : | : | :
Db 1006 ELAAEKEKISVLQNNYEKSOETFQLQDSDFYGRESSELLATRDQLKSVEEKL S-----LAQ 1060
| : | : | : | : | : | : | : | : | : | : | : | : | :
QY 977 KDMIT-----WILQMFDTYVKRLREKS-----CLIVAITSSESINKDNFIT 1016
| : | : | : | : | : | : | : | : | : | : | : | : | :
Db 1061 EDLISRNRIGNQMKLQELTKATLEQDSAKKEQOLBRKCALODIQEKKALEKELV 1120
| : | : | : | : | : | : | : | : | : | : | : | : | :

Qy 1017 HDK-----EKAERKRKAEARLHRQKIMQMSALQK--NFIETHKLIMYDNTSEMPCKD 1068
Db 1121 NEKSLAEIEIKRQSEKTKTKNEELKSHKLESIKETITNLKDAKQLLQKBELOQKAD 1180
Qy 1069 SI----MEESTPAVSRYRALTGPKRGPSTVEVLTCILCQEQEVK----IENNAWVLS 1122
Db 1181 SLKAAVEQE-----KRNOQLKQDVK-----KKEEELKFEIEKAKLHS 1220
Qy 1123 ACVQKSTALTOHRGKPIELS 1142
Db 1221 EIREKEVGMKKHEENAKLT 1240

RESULT 15
T08621
centrosome associated protein CEP250 - human
C:Species: Homo sapiens (man)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: T08621
R:MacK, G.J.; Rees, J.; Sandblom, O.; Balczon, R.; Fritzler, M.J.; Rattner, J.B.
Arthritis Rheum. 41, 551-558, 1998
A:Title: Autoantibodies to a group of centrosomal proteins in human autoimmune sera
A:Reference number: 216462; MUID:98165428; PMID:9506584
A:Accession: T08621
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2442 <MAC>
A:Cross-references: EMBL:AF022655; NID:g2832236; PIDN:AAC06349.1; PID:g2832237
A:Experimental source: cell line HeLa

Query Match 2.0%; Score 182; DB 2; Length 2442;
Best Local Similarity 20.9%; Pred. No. 0.019;
Matches 256; Conservative 187; Mismatches 478; Indels 304; Gaps 60;

Qy 212 WE--EKKELPELOIRKKNRYCYLVNDEHSHYDVHVIYSLQRALDCELAE-----AQ 262
Db 88 WENVEEPNL-DELLVRLLEEQQRCESLAE-----VNTQIRLHMEKADVVNKKALRAD 137
Qy 263 LHTTAIDKEGRPAVKAGAYACQAEKEDIKSHSENVSOHPLHVEVLHSEIMAHQKFARLR 322
Db 138 VEKUTVDMSRADELMKESQWQMEQEFFKGKLGEGHRLLS---LWREVVTFRRHFLFM 194
Qy 323 GSWMNMK-IMSYSSDFROI---FCOACLR-----EEP-----DSENPCLISRLMLWDA 365
Db 195 KSATDRDLMELKAEHVRSLGSLTCLCLTLVTAQSRPENGSGRMDGREPAQLLLLLAKTK 254
Qy 366 KLYKGARKILHELIFSFFMEMEYKKL--FAMEFVKYKQLOKEYISDDHRSISITALS 423
Db 255 ELEKEAHERSQELTQKSGDLEKAEQLQDRVTELSALLTOSQK--NEDYEKMIKALRET 312
Qy 424 VQMTFV-----PILARHLIE---QNVISVITETLLEV-----LPXYLDNRN 461
Db 313 VEILTNHTLMEHAESLSRNAQEKLSQLQGVKIDITQVMVEEGDNTAQSGSLNLELE 372
Qy 462 ---NKTNFOGYSQDK-----LGRVYAVICDLKY-----ILISKPTIWE-- 497
Db 373 SSIFSQFDYQ--DADKALTLYRSVLTTRRQAVQDLRQOLACQCAVNLQOQHDQWEEG 430
Qy 498 ---RLRMQLEGFRSFLKILTC-MQG-----MEEIRVOVGQHIEV--DPDWE-A 539
Db 431 KALQRQLKLTGERTLAGQVTDLQGEVDSLSKERELLQKARELRQOLEVLEQEAWRLR 490
Qy 540 AIATOMQLKNTLLMPOWCACDEBLLVAYKECHKAVYKRCSTSFSSKTVVQSCGHSLE 599
Db 491 RVNVELQIQGDSAGQK-----EE-----QOEELHLAV-----RERERLQBMGLM 532
Qy 600 TKS-----YRVSGEDLVSIHLP---LSRTLGLHVRSLRGAVSRLHEFVSFED-FOVE 648
Db 533 AKQGESUSELITLREALSHLEBELRLQEQOTEVTAMALAR--AQSTAEILSSSNTLKTE 590
Qy 649 VLVETPRLCLVLVAQVVAEMRRNGLSLISQVFYY-----QDVKREEMYDKDIIMLQIG- 703
Db 591 VA---DLRAAAVKLSALNEALALQKVLGNLOLLQLEBENQSVCSMRMAAEQARNALQVDL 647

QY 704 -----ASIMDPNKELLVLORYELA-----EAFNKTIKQODLIKQYN----- 742
Db 648 ABAEKREALWEKNTHLEAQLQAEAGAEALQADLQIOEKEEIQKIKSSESRHQEPAAT 707
QY 743 TLIEEMLQ-----VLIIVIGERYVPGVGNVTKKEEVTWREIHHLLCIE-----PMP 787
Db 708 TOLEQLHQEAKRQEEVLARAVQEKAE-----LVREKAAL-VRLOAVERDRODLAAQLQ 760
QY 788 HSAIAKNLPENNETGLN-VIN-----KVATFKPGVSGHYVELKDESLKDF 836
Db 761 GLSSAKELLESSLFEAQOQNSVIDEPQGLEQIQVTOAKEVIOGE-VRCLKLE----- 814
QY 837 NMYFYHYSTQHSKAHMKRKRQENKDEALPPPPPEFCPAFSKVINLLNCDIMYIL 896
Db 815 -----LDTERSQAEOERDAARQLAQAEQEGKTALQEQKAAHEKVNQ-----L 858
QY 897 RTVFERAIDTDSNLTWTEGLMAFHILALGLLEKQOQLOKAPPEEVTDFYHKASRLGSS 956
Db 859 REKWEK-----ERSWHQOELAKALESLEREKMELEMLRLKEOQTEMEAIQAREERTQAE 913
QY 957 ANNIQMLEKLGIPQLEGQKDMITWILQMFDTVKRLREKSCLI--VATTSGSESINKDE 1014
Db 914 SALCOM-----QLETEKERVSVLLETLIQTKELADASQOLERLRQDMKVQKLKEQE 964
QY 1015 IT-----HDKEKAERKKAEARLHROKIMA---OMSALQKNFIETHKLMYDNTSEMPGKE 1067
Db 965 TTGILQTLQEAQRELK-EAARQHRDDLAALQAESESSLLQDKMDLQKQVEDLKSQVAD 1023
QY 1068 DS--IME--EESTPAVSYSRIALGPKRGPSVTEKEVLTCLCQEEQEVKIENNAMVLS 1122
Db 1024 DSQRLVEQVQKRLRETQYNRIOKELER-----EKASLTSLMEKEQRLLVLOEAD--S 1076
QY 1123 ACVQKSTALTOH-----RGKPIELSGE-----ALDPLFMDPDLAYGTYTSGCGHV 1167
Db 1077 IRQOELSAURQDMQEAQGEQKELSAQOMELLRQEVKEKEADFLAQEAQLLEEL--EASHI 1133
QY 1168 ----MHAVCWQKYFEAVQLSSQORIHVDLFDLESGEYLCPLCKSLCNTVPIPIPOFOKI 1223
Db 1134 TEQQLRASLWQAEAKAAQL-----HLRLRSTES-----QLEALAAEQ 1171
QY 1224 NSENADALAOILLTARWIOTVLARI 1248
Db 1172 PCNQAAQALASLYSALQOALGSV 1196

Search completed: September 25, 2003, 14:56:30
Job time : 78 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 25, 2003, 14:46:05 ; Search time 33 Seconds
(without alignments)
2471.040 Million cell updates/sec

Title: US-09-724-126a-19

Perfect score: 9141

Sequence: 1 AMEGNMADEAGGTERMEIS.....ETASQETNQMLFGFNQQL 1734

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	990	10.8	1958	UBR1_SCHPO	O60152 schizosacch
2	866	9.5	2052	UBR8_SCHPO	O13731 schizosacch
3	691.5	7.6	1950	UBR1_YEAST	P19812 saccharomyc
4	628.5	6.9	1941	UBR1_KLULA	O60014 kluyveromyc
5	208	2.3	2230	G064_HUMAN	Q13439 homo sapien
6	190.5	2.1	3259	G1AN_HUMAN	Q14789 homo sapien
7	189.5	2.1	8797	SNE1_HUMAN	Q8nf91 homo sapien
8	182.5	2.0	2442	CEP2_HUMAN	Q9bv73 homo sapien
9	177	1.9	6885	SNE2_HUMAN	Q8wxh0 homo sapien
10	176	1.9	1829	MY3A_CHICK	Q02440 gallus gall
11	175.5	1.9	1427	REST_HUMAN	P30622 homo sapien
12	175.5	1.9	1526	MY52_SCHPO	Q9usi6 schizosacch
13	175.5	1.9	2663	CENE_HUMAN	Q02224 homo sapien
14	170	1.9	2298	CU05_HUMAN	Q9y3r5 homo sapien
15	168.5	1.8	3911	AK9A_HUMAN	Q99996 h a-kinase
16	166	1.8	1163	SBCC_CLOAB	Q97fk1 clostridium
17	164	1.8	2096	BP28_DROME	Q9vm75 drosophila
18	162	1.8	2335	TOR1_SCHPO	O14356 schizosacch
19	158	1.7	2033	EVPL_HUMAN	Q02817 homo sapien
20	157.5	1.7	3214	BPAL_HUMAN	Q03001 homo sapien
21	156.5	1.7	1790	USO1_YEAST	P25386 saccharomyc
22	156.5	1.7	2710	TOXA_CLODI	P16154 clostridium
23	156.5	1.7	2869	RBPL_PLAVB	Q00798 plasmodium
24	156	1.7	1853	MY5A_MOUSE	Q99104 mus musculus
25	155.5	1.7	1216	P1B1_RAT	P10687 rattus norv
26	155.5	1.7	8545	ANCL_CAEEL	Q9n4m4 caenorhabdi
27	154	1.7	2469	TEGU_HSVSA	Q10156 herpesviru
28	154	1.7	3685	DMD_HUMAN	P11532 homo sapien
29	153.5	1.7	1875	MLP1_YEAST	Q02455 saccharomyc
30	153	1.7	1682	MSPL_PLAF3	P15998 plasmodium
31	152.5	1.7	1325	G160_MOUSE	P59397 mus musculus
32	152.5	1.7	3433	UTRO_HUMAN	P46939 homo sapien
33	151.5	1.7	1855	MY5A_HUMAN	Q9y411 homo sapien

ALIGNMENTS

RESULT 1

ID	UBR1_SCHPO	STANDARD;	PRT; 1958 AA.
AC	O60152;		
DT	15-DEC-1998 (Rel. 37, Created)		
DT	15-DEC-1998 (Rel. 37, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	N-end-recognizing protein (Ubiquitin-protein ligase E3 component) (N-recognin).		
GN	UBR1 OR SPBC19C7.02 OR SPBC32F12.14.		
OS	Schizosaccharomyces pombe (Fission yeast).		
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;		
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;		
OC	Schizosaccharomycetes.		
OX	NCBI_TaxID=4896;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=21559218; PubMed=11702950;		
RA	Kitamura K., Katayama S., Dhut S., Sato M., Watanabe Y., Yamamoto M.,		
RA	Toda T.;		
RT	"Phosphorylation of Mei2 and Stell by Pat1 kinase inhibits sexual		
RT	differentiation via ubiquitin proteolysis and 14-3-3 protein in		
RT	fission yeast."		
RL	Dev. Cell 1:389-399(2001).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=972;		
RX	MEDLINE=21848401; PubMed=11859360;		
RA	Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,		
RA	SGouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,		
RA	Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,		
RA	Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,		
RA	Gentles S., Gobie A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,		
RA	Holroyd S., Hornsby T., Howarth S., Huckie E.J., Hunt S., Jagels K.,		
RA	James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,		
RA	Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,		
RA	Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,		
RA	Rutherford J., Rutter S., Saunders D., Seeger K., Sharp S.,		
RA	Skellton J., Simmonds M., Squares R., Squares K., Stevens K.,		
RA	Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,		
RA	Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,		
RA	Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,		
RA	Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,		
RA	Borzyn K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,		
RA	Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,		
RA	Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,		
RA	Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,		
RA	Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,		
RA	Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,		
RA	Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,		
RA	Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,		
RA	Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;		
RT	"The genome sequence of Schizosaccharomyces pombe."		
RL	Nature 415:871-880(2002).		
CC	-1- FUNCTION: RECOGNITION COMPONENT OF THE N-END RULE PATHWAY. BINDS		
CC	TO PROTEINS BEARING AMINO-TERMINAL RESIDUES THAT ARE DESTABILIZING		

P11533 gallus gall
Q15413 homo sapien
Q9upn3 homo sapien
Q15078 homo sapien
Q9gkx0 mus musculu
Q96pk2 homo sapien
O67124 aquifex aeo
Q9tu23 bos taurus
P40457 saccharomyc
Q99323 drosophila
P32874 saccharomyc
Q9utk5 schizosacch

Db 683 NNNATLLTOSNRSTL-----FSSDIETPTIQLNRQVLKTRRTNLSFSDLG 732
QY 486 YIL-----ISKPTIWTIRLMOLEGRSFLKILTCMQGMEIRROVQGHIEPDDWEAAI 541
Db 733 YLLQHPQVKKLVVDDFRYVHQYIDLLRVE-----QGVIPQORAILSHVQ-----WDFPH 781
QY 542 ATOMQLKNILLPOEWACADEL-----LLVAYKECHKAVNRCSSTFISSSKTVVQSGGHS 597
Db 782 G-----KNILFVNRVAMLSNTVSSCFTQAPYERLEYAI-KCIIITSITHPKLDI-----AES 832
QY 598 LETKS-----YRVSDLVSIHLPLSRTLAGLHVRLSRLGAVSRLHEFVS 641
Db 833 LEPLSCIPSSLTNTQPLVPFSVRDPIISFYHP-----LHMLSNLSFYCRVDASSH 885
QY 642 FEDFQVEVLVEXPLRCLVLVAQVAMRRNGLSISVYFQVQVDRCREEMYDKDIIMLQ 701
Db 886 WDKDTLLALLDPLRVCVLLAQIDCNLWRNGRSILLTDAFYQQLNIEVSYDKDILAIQ 945
QY 702 IGASLMDPNKFLLLVLRQYELAE-APNKYISTKQDOLIKOYNLTIEMLQVLIYVGERY 760
Db 946 TILMFDPNVLNAVQVREFTDNLNLTYNNEHPNYDTERIPAMLCMKLELLIALITER- 1004
QY 761 VPGVGNVTRKEEVTMRIIHLCTEPMPSHAIAKNLPENNENETGLENVINKVATEKKP-G 819
Db 1005 -EQILHVDIODIIRTLAOLQCGPLAYSALLSTISSNLVESLSFKIREVTSYKAPDG 1063
QY 820 VSHGVGYELDESKDFNMYFYHYSTQHSKAEHMOKRRKQBN--KDEALPPPPPEFC 877
Db 1064 LHDGFGYSLKDEYDLVDPYFYHYNKNERESTILKRLAKKNVSAESIIEPKIRFL 1123
QY 878 -----PAFSKVINLLNCDIMMYILRTVFERAIDTNSLWTEGMLQ-APHILALG--LLE 929
Db 1124 EKDGHDIFFAVNAFTSLIIF-----RAIE-----YALVQAESFGSDIGNITLG 1169
QY 930 EKQOL-----QKAPEEVEVTFDYHKA-----SRLGSSAMNIQMLELKGIPQL 973
Db 1170 DALQCLISMKIHEFSKSNDFCSRAERYPTDSSIMREFGGAYCAELCFAILKSPKY 1229
QY 974 EGOKDMITWI---LQMFDD--TVKRLREKSLIIVATTSGSESIRKDEITHDKAEKRRKA 1028
Db 1230 KDVHVKNVAVLAGLQKNDPSAYSNMLEATHFELSTTSS-----TSDSNELEKTQEK 1280
QY 1029 EAARLHQ-KIMQMSALOKNFETHKLMYDNTSEMPGK---EDSIMEESTAVSDYSR 1084
Db 1281 KRLALEKQKIMQOFRDOQASFLA-----QNTDFDGEDQTEDEVTEPEEVEKYHEH 1334
QY 1085 IALGPKRGPVSVEKVLCTICQEOEVK-----IENNAVLSACVOKSTALTOHRG 1136
Db 1335 I-----RG-----NCLLCOECNDQAPYGVIGLIQSSLLRKTVDHSEIILDEIYS 1380
QY 1137 KPIELS-----GEALDPLFMDPD-----LAY-----GYTGCSCGHVMHACVQKY 1176
Db 1381 VPNLDRESHSRPFGKKYDVTNVFNRSKDRLLSAYPPGNNIRGVFVSCGHLMLHLCFKNY 1440
QY 1177 FEAVQLSSQORIHVDFLESG--EY-----LCPLCKSLCNTVPIIPLQPOKIN 1224
Db 1441 YVARSMYRN-----DVTAGLSEYKYKSTAKFMCPICRSILSNVLLP-MQOIPKM-- 1489
QY 1225 SENADALAQLLTLARWIOTVLARISGYNIRHAGENPIPIFFNQGMDSTLEPHSI----- 1280
Db 1490 CLNIDTLNFRSMNGWLEET-GTMSSSSEFYQ-----LVRSSLSDDTKDTRFSCFLRP 1540
QY 1281 -----LSFGVESSIKYSN--IKEMVILFATTIYRIGLKVPPDERPRV-PM 1324
Db 1541 WINSKIISAMLAELKATDAGALIDQSNRNVSDIYDRYCETT-KLAMKLKLVGSGFTNVSPH 1599
QY 1325 LTWSTCAFTIQTATENLLGDEGPK-LFGALQNRHNGLKA-MOFAVAQRITCQVLIQKH 1382
Db 1600 DLLNSLAYTVSSLE--VSQRCSPKQSGATRSVWFNGLPLTLFPLTSLD----- 1648
QY 1383 LVRLSLVPLNTKSEDTPCLLSTDLFLHVLGAVLAPPSLY-----WDDPVDLQPSV 1434

Db 1649 ---VLKVCVDIOIKSDQOALLMESQKLLVCKIFRHSQKSLMRLNGRMSDHDIOQPELL 1705
QY 1435 SSSY-----NHLYLPHLITMAHMLQLLITVDTGLPLAQVOED---SEE 1474
Db 1706 SNTFDDFVKISSMLIFGKQDNILYVVKLFSEICKTIISMIVKVVADSSVVPDLTINYS 1765
QY 1475 AHSASSFAELISQYTSIGCDIPGWYLWLSKNGI-----TPYL 1514
Db 1766 QQSKSQFY-----ILC--KNVLLWCSSNNIETLDDDESLLRLMSLVKYSPLPFL 1813
QY 1515 RCALAFHYLILGVTPP-EELHTNSAEGYSALCSYLSLPTNLFLLFQEWY-----DT 1565
Db 1814 RRVALVLYCHPDISLEFNEFSNEDDSLERLSKLIKVP-----LQELYSSOMSDENQ 1868
QY 1566 VRPLLQRCADPALLNCLKQNT-----VVRYPRKRNLSIE-LPDDYSCL 1609
Db 1869 ILEIAGWCEHLA-----QNTWGDSTISLEYPGIVELVPLHRLLENLIDSMQESVCCM 1921
QY 1610 LNOASHFRCPSRADDERKHPVLCFLCGAILC---SONICCOEIVNGEEVGACIFHALHC- 1665
Db 1922 CHKTPIIL-----PAICMLCGSVICFNARQNTVSSRRLTGE---CNKHAATCT 1965
QY 1666 -----KA-----RGCAVPAPYLDEYGETDGLKRGNPLHLRSRYR-KLH 1704
Db 1966 GSVCIFFITKACGILLLDSISNTGTIMPTPYLDIHGETDQLRRCGPOFLNOKRYDFVVR 2025
QY 1705 LVMOQHCHIEIARSQETNQLMFGFNOL 1733
Db 2026 EOWLRQTVLQKMARHMDMTEMQ---NWRM 2051

RESULT 3

UBRL_YEAST
ID UBRL_YEAST STANDARD; PRT; 1950 AA.
AC P19812;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DE 15-JUL-1998 (Rel. 36, Last annotation update)
DE N-end-recognizing protein (Ubiquitin-protein ligase E3 component) (N-
DE recognin).
DN UBRL OR PRT1 OR YGRI84C OR G7168.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / GRF88;
RX MEDLINE=91006011; PubMed=2209542;
RA Bartel B., Wuening I., Varshavsky A.;
RT "The recognition component of the N-end rule pathway";
RL EMBO J. 9:3179-3189(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=97279231; PubMed=9133739;
RA Arroyo J., Garcia-Gonzalez M., Garcia-Saez M.I., Sanchez-Perez M.,
RA Nombela C.;
RT "DNA sequence analysis of a 23,002 bp DNA fragment of the right arm
RT of Saccharomyces cerevisiae chromosome VII";
RL Yeast 13:357-363(1997).
RN [3]
RP SEQUENCE OF 346-1950 FROM N.A.
RA Hebling U., Hofmann B., Delius H.;
RA Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: RECOGNITION COMPONENT OF THE N-END RULE PATHWAY. BINDS
CC TO PROTEINS BEARING AMINO-TERMINAL RESIDUES THAT ARE DESTABILIZING
CC ACCORDING TO THE N-END RULE, BUT DOES NOT BIND TO OTHERWISE
CC IDENTICAL PROTEINS BEARING STABILIZING AMINO-TERMINAL RESIDUES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its


```
QY 1586 KNTVRYPRKRN-----SLLEPDDYSCLLNQASHRCPRSD-----DERKHPV 1630
Db 1642 QESIVSHPYLENIPYECGIIKIDLSKYLNTYVYQSKIEIKLREERSOHMKNAONRLDFK 1701
QY 1631 LCLFCGAIL-----SQNICCCQIVNGEEVGCIFHALHCKARGCAYPA 1674
Db 1702 ICLTCGVKVLRADRHETMKLNKNCFKPGCAFMPNSSEV-CLH-LTQPPSNIFISA 1757
QY 1675 PYLDEYGETD-PGLKRNPHLSRERYKRLHLVWQHCIIIEIARSQE-----NQM 1725
Db 1758 PYLNSHGVEGRNMRGDLATLNLKRYEHLNRLWINNEIPGYSIRVMGDEPRVTILSNFG 1817
QY 1726 LFGFN 1730
Db 1818 LFAFN 1822

RESULT 4
UBRL_KLUJA STANDARD; PRT; 1941 AA.
AC O60014;
DT 15-DEC-1998 (Rel. 37, Last sequence)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE N-end-recognizing protein (Ubiquitin-protein ligase B3 component) (N-
GN recognin).
OS Kluyveromyces lactis (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
OX NCBI_TaxID=28985;
RN [1]
RP SEQUENCE FROM N.A.
RA Waller P.R.H., Varshavsky A.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: RECOGNITION COMPONENT OF THE N-END RULE PATHWAY. BINDS
CC TO PROTEINS BEARING AMINO-TERMINAL RESIDUES THAT ARE DESTABILIZING
CC ACCORDING TO THE N-END RULE, BUT DOES NOT BIND TO OTHERWISE
CC IDENTICAL PROTEINS BEARING STABILIZING AMINO-TERMINAL RESIDUES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF061554; AAC15841.1; -
DR PIR; T30554;
DR InterPro; IPR003126; Znf_Nrecognin.
DR Pfam; PF02207; zf-UBR1; 1.
DR SMART; SM00184; RING; 1.
DR SMART; SM00396; Znf_UBR1; 1.
KW Ligase; Ubl conjugation pathway.
SQ SEQUENCE 1941 AA; 223682 NW; 37C2E1BCA0803268 CRC64;
```

```
Query Match 6.9%; Score 628.5; DB 1; Length 1941;
Best Local Similarity 20.4%; Pred. No. 2.7e-30;
Matches 406; Conservative 315; Mismatches 683; Indels 587; Gaps 92;

QY 40 FYTAFLLHLAQLVPEIYFAEMDPLEKEESVQMSIETPLEWYLFGEPPDICLEKLHSG 99
Db 51 YYYNIMSDSGRLPHMTATNAREFPNVDQAMEIKLSKPKYKIDENG--HGFNFHAG 108
QY 100 AFQLCGVFKSGEITYSCRCADIPTCVLCMDCPQDSVHKNHRY--RMHTSTGGGFCDCG 157
Db 109 --RICGAKFVGEPYIRCKESFDDTCVLCVNCNPNKPDHGVGHVYTSICTFENGICDCG 166
QY 158 DTEAWTKGPCVNCHEPGRAGTIKENSR-----CPLNEEVIQVARKIIPSPVIKY 205
Db 158 DTEAWTKGPCVNCHEPGRAGTIKENSR-----CPLNEEVIQVARKIIPSPVIKY 205
```

```
Db 167 DKEAW-----NHENLCKGA-EDNGRLEDEDDHDKISKMLSVLIELFDHFDIVFNQ 218
QY 206 VVE--MTWE-----BEKELPPELQIREK-----NER----- 230
Db 219 NIEPLTTIQKPLIAKRYFNPEREYEQADMLRLRAYRNQYMDDESSNKRHLTSDPLS 278
QY 231 ---YVCVFNDEHSHYVHSIORALDCELAQLHTTADKEGRRVAKAGA-YAACQ 285
Db 279 TLKDYAILVYDEFYNSQASAAIRQG-GPDKNKHIDLLTAKIDSEGRSLRCSADIASLM 337
QY 286 EAKEDIKSH--SENYSQHPHVEVLRSEITMAHQFALRLGSGWMNKIMSY-SSDFRQIFCQ 342
Db 338 GRIFSQVSNGLSCTITQ--WYELHQEAC---KYSI---MWINDCLNPSTFQSLFRN 368
QY 343 A-----CLREEDP-----SENPCLI----- 357
Db 389 AIGKVLCSKYEPFYQSIDMTSVVRDYFSDSYLSDDPYLAHDHVLGEGVKPIGRHKSLD 448
QY 358 -----SRL---MLWDAKLYKGARKILHELHIFSSFFMEMEYK 390
Db 449 PGDISAISPLNKNKVIADHHEYNTRLQYVLFLENRYWKKRKIVODLIPTLASSAVOK 508
QY 391 KLFAFEFVKYKQLOKEIYSDHDSISITALS---VOMFTVPTLARHLIEE-----Q 440
Db 509 PMFTDQVLEIFPHMTRS--GTFMDREPOLTSRESVQLFTCTPTTAYSIFHSGHFNLIW 566
QY 441 NVISVITE--TLLE--VLPEYLDNRNKNFQOYSODKLGVRVAVICDLKYI---LISKP 492
Db 567 SVIDVDFSTMDGTLVMQVRQSRNP--SKSYSISFKQGLYAVETLLSKITDNPULLKP 624
QY 493 TIWTERLMQFLEGFRSLKILT---CMOGMEI--RRQVGOHI-EVDPDWEAAIAIOMQL 547
Db 625 -----GEFIMIVTCLKFNAGWKIKKKREGEHVLRDQHFIPYLEYTTVS 669
QY 548 KNILLMF-----QEWACACDEELLVAYKECHKAVMRCSTSFSSSTVTVQSGCH----- 596
Db 670 YSIITFTKVLQOQSKDHIDQRLIGA-----INLLDSFL-----GHRNLSYK 711
QY 597 ---SLETSKYRVSDELVSITHPLSRTLGL--HVRLS-RLGAVSRILHEVSEDFQVEVL 650
Db 712 LYKDEPIIKFKQISKEQVSPMNPVHTLFCFLVQHPVLPQSVLQSVLSQSKDLYLISDF 766
QY 651 VEYPLRCLVLVAQVAVEMWRRNGLSLISQVFFYQDVVKCREM--YDKDIIMLIQIGASLMD 708
Db 767 ---ALRSVVLCSQIDIGFWRNGMSVLIHOSAYY---KNPEMSSYSRDLQNLQ-AFLIE 819
QY 709 PNKFLLV---LQRYELAEAFNKTIQDQDLIKQYNTLIEEMQVLVIYVGRYVPGVG 765
Db 820 KNDFORVIYNMLDRWELLDWFDGSPSTETVYDDKISSIIQQFVAFLYQILVER----- 873
QY 766 NVTKEEVTMRE-----IIHLGIEPMPHSAIAKNLPEN-ENNETGLENVINKVATF 815
Db 874 DFYKKFDTLEETQLYNIKNAIIVKLYAEPLSVTDLNDIPDLVLTESVSQFDTVLEEVSTY 933
QY 816 KPGVSGHGVYELKDESLKD-----FNM--VFYHYSKTOHKSAAEMQKRRKQENKDEA 867
Db 934 IEPKSGKIWCFTQKEGTLOENRFLRLNMGNDFEHSATIVKSHLADSCKRAKIIYKPK- 992
QY 868 LPPPPPPPECFPAFSKVINLLNCDIMMYILRTVFERAIDTDSNLMTGEMQLQMAFHILALGL 927
Db 993 -----LLELDLDPCCAR---ELGSFTRTNLFA-----KLIFKLKULAV 1027
QY 928 LEEKQOLQKAPEEVTFDFYHKASRLGSSAMNQLLEKLKGPQLEQOGKDMITWILQMF 987
Db 1028 -----SDSSFSTYELLHLIHAIFRDEENVGKDSLPEAVISKPICDLLLLSIV 1075
QY 988 DTVKRLREKSCULIVATTS-----GSESIKNDEITH-----DKE 1020
Db 1076 DSEGSFSFSEN--VWATADYLLDNMIMKRPATAVLESLETCFGTKYIADYKIRKANQGVNFE 1133
QY 1021 KAERYKAAEARLHRQK IMAQMSALOKNFIEHTKLMYDNTSEMPGKEDSIMEESPAVS 1080
Db 1134 ETEQERKRRLAKNRQOOIMNRFSSROPKFMDKHE-EYSAGND-----EDVDMOGEDLAG-- 1186
```

```
QY 1081 DYSRIALGPKRGSVTEKEVLTICLQEEQEVKIENNAMVLSACVQKSTALTQHRKPIE 1140
Db 1187 -----ELNEFHGSLCHDD-----VSDDFFVPIPIYQNVSPVFLSNPTPME 1226
QY 1141 L-----SGEALDPLFMDPLDAYTYTG-----SCGHVHAWCQKQYF 1177
Db 1227 IYKPHGFGFNNHLL--ATYNTDLFYKKKENGASQLMHSTQKVLVSCNHAHVYRCFKHYI 1284
QY 1178 EAVQLSSQRIHVDLFDLSEGYELCPKLSCLNTVPIPIPLQPKQINSENADALAQLTL 1237
Db 1285 D-----KRYSTDLF-----ICPLCQTCYNSVIPV-----DTVKLQSGDRLLQOKLT 1326
QY 1238 ARWIOTVLARISGYNIRHAKGENPIPIFFNOGMDSTLEPHSILSGVESSIKVYSIKE 1297
Db 1327 GGLDSLLTFEYSSSE-----CNDEVGKIIL-----SLKDSNN--- 1360
QY 1298 MVILFATTIYRIGLKVPDPDRPVMPLTWSTCAFTQATF-----NLLDGEGRPLF 1349
Db 1361 -----GLRL--NRNDP-----TWIQDRFLTSLQFSNNICLLEMLSLNKNKDPF 1401
QY 1350 GALQNRQHNGLKALM-----OFAVAQRITCPQVLIQKH-----LVRLL 1387
Db 1402 GTLLSGEQKFTQLNKLASLAVYTRLT-----KHEMYSNSMKIFVAAIYQVIRFP 1454
QY 1388 SVVLPNIKSEDP--PCLLSIDLHVLVGAVLAPPSLY--WDDPVDLQPSVSSYNHLYL 1443
Db 1455 RVETVLRSLRSLFKDCLQEV--LVERLKLTKDQGSFYRYESELRAQTCLDSSEFS--- 1509
QY 1444 PHLITMAHMLQILLTVDTGLPLAQVQEDSEEAHSASSFAEISQYTSGSGICDIPGWLW 1503
Db 1510 -----IVLTKTILAGRG-----DQVEKHTLDLFY----- 1534
QY 1504 VSLKNGITPLRCAALFFHYLLG-VTPPELHTNSAEGEYSALCSYLSPTNLFLLFQY 1562
Db 1535 TELISELLPTLRSIILLKALKQFMGTGGDDDFNEKDVLSGSLASE-SREKHFHLL--- 1589
QY 1563 WTVRPLRQRCADPALLNC---LKOKNTVVRPKRNSLIELPD---DYSCLLNQASHF 1616
Db 1590 ---IRFLQOTDYD--LLMNSHSPLESSELVNAHPDHCYSIIKTLDTLNTVYTNKNI 1645
QY 1617 RCPRSADDERKHVP-----LCLECGAILCSQNCICQEVNGEVGACIFHALHCK--- 1666
Db 1646 TLREENDQKIRTVNRLVDYKICLCGVKTHAR-----TDGLEMOK---HMERCSHGSS 1695
QY 1667 -----ARGCAY-----PAPYLDXYGETD-PGLKGRGNPLHLRERYRKLHLVW 1707
Db 1696 GFLIPNISQVCLYLSRPDCTVNISAPYLSHGSGRNAIERGDLTVLNHARYEHLTRLW 1755
QY 1708 QQHCHIEETAR 1718
Db 1756 ISNGIPGYISR 1766
```

RESULT 5

```
GOG4_HUMAN
ID GOG4_HUMAN STANDARD; PRT; 2230 AA.
AC AC13439; Q13270; Q13654; Q14436;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Golgi autoantigen, golgin subfamily A 4 (Trans-Golgi p230) (256 kDa
DE golgin) (Golgin-245) (72.1 protein).
GN GOLGA4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=96215236; PubMed=8626529;
RA Erlich R., Gleeson P.A., Campbell P., Dietzsch E., Toh B.-H.;
RT "Molecular characterization of trans-Golgi p230: a human peripheral
```

```
RT membrane protein encoded by a gene on chromosome 6p12-22 contains
RT extensive coiled-coil alpha-helical domains and a granin motif.";
RL J. Biol. Chem. 271:8328-8337(1996).
RP SEQUENCE FROM N.A.
RA Seelig H.P.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 131-2230 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=96125112; PubMed=8537393;
RA Fritzler M.J., Lung C.-C., Hamel J.C., Griffith K.J., Chan E.K.L.;
RT "Molecular characterization of golgin-245, a novel Golgi complex
RT protein containing a granin signature.";
RL J. Biol. Chem. 270:31262-31268(1995).
RN [4]
RP SEQUENCE OF 524-672 FROM N.A.
RC TISSUE=Gastric fundus;
RA Balague C.;
RL Thesis (1994), Instituto municipal de investigacion medica, Spain.
CC -!- FUNCTION: May play a role in vesicular transport from the trans-
CC Golgi.
CC -!- SUBCELLULAR LOCATION: ASSOCIATED WITH THE CYTOPLASMIC FACE OF THE
CC GOLGI MEMBRANE.
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=4;
CC Comment=Additional isoforms seem to exist;
CC Name=1;
CC IsoId=Q13439-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q13439-2; Sequence=VSP_004272, VSP_004273;
CC Name=3;
CC IsoId=Q13439-3; Sequence=VSP_004274;
CC Name=4;
CC IsoId=Q13439-4; Sequence=VSP_004275;
CC -!- DISEASE: ANTIGEN IN THE AUTOIMMUNE DISEASE SJOREGREN'S SYNDROME AND
CC IN HEPATITIS B.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U41740; AAC50434.1; -
DR EMBL; X82834; CAA58041.1; -
DR EMBL; U31906; AAC51791.1; -
DR EMBL; X76942; CAA54261.1; -
DR Genew; HGNC:4427; GOLGA4.
DR MIM; 602509; -
DR GO; GO:0005802; C:Golgi trans-face; TAS.
DR GO; GO:0016192; P:vesicle-mediated transport; TAS.
DR InterPro; IPR000237; GRIP_domain.
DR Pfam; PF01465; GRIP; 1.
KW Golgi stack; Antigen; Coiled
KW DOMAIN 133 237
FT COILED COIL (POTENTIAL).
FT COILED COIL (POTENTIAL).
FT COILED COIL (POTENTIAL).
FT COILED COIL (POTENTIAL).
FT COILED COIL (POTENTIAL).
FT TPYKGNLYHTDVSLSFGPTEPEYLRKVLFEY -> HLTKV
FT AICTIRMSHLENPLNSICEKFLSI (in isoform
FT 2).
FT /FTId=VSP_004272.
FT Missing (in isoform 2).
FT /FTId=VSP_004273.
FT Missing (in isoform 3).
FT /FTId=VSP_004274.
FT FTSPRSGLF -> SHLRSSS (in isoform 4).
FT /FTId=VSP_004275.
FT Y -> K (IN REF. 3).
FT R -> H (IN REF. 3).
FT Y -> H (IN REF. 3).
FT CONFLICT 188 188
FT CONFLICT 220 220
```

Db 1291 SFOATHQLLEKENQIKSMKADIESLVEKEALQEGGNOQQAA-----SEKS 1333

Qy 998 CLIVATTSGSESINKDEITHDKKAERKRKAEEAARLRQ-----KTMAQM 1042
 |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Db 1340 CITQLKKELSEN--NAVTLMKLEEK-KVEISSLSKLDTLNVLONSISLSREAAI 1396

Qy 1043 SALQKNFIEHTHKLWYDNTSMPCKEDSIMEEE--STPAVSYS 1083
 |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Db 1397 SSLRKQVDEBKCELLDQVDLFSPKVDTLSKEKISALESQVDWS 1439

RESULT 6
GIAN_HUMAN
ID GIAN_HUMAN STANDARD; PRT; 3259 AA.
AC Q14789; O14398;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Golgi autoantigen, golgin subfamily B member 1 (Giantin) (Macrogolgtn)
DE DE Golgi complex-associated protein, 372-kDa) (GCP372).
OS GOLGB1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_Taxid=9606;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94187728; PubMed=7511208;
RA Seelig H.P., Schranz P., Schroeter H., Wiemann C., Griffiths G.,
RA Renz M.;
RT "Molecular genetic analyses of a 376-kilodalton Golgi complex membrane
RL protein (giantin).";
RL Mol. Cell. Biol. 14:2564-2576(1994).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94257116; PubMed=8198703;
RA Seelig H.P., Schranz P., Schroeter H., Wiemann C., Griffiths G.,
RA Renz M.;
RT Macrogolgin -- a new 376 kD Golgi complex outer membrane protein as
RT target of antibodies in patients with rheumatic diseases and HIV
RT infections.;
RL J. Autoimmun. 7:67-91(1994).
[3]
RP SEQUENCE FROM N.A.
RX MEDLINE=95100974; PubMed=7802676;
RA Sedha M., Misumi Y., Fujiwara T., Nishioka M., Ikehara Y.;
RT "Molecular cloning and sequence analysis of a human 372-kDa protein
RL localized in the Golgi complex.";
RL Biochem. Biophys. Res. Commun. 205:1399-1408(1994).
CC -!- FUNCTION: May participate in forming intercisternal cross-bridges
CC of the Golgi complex.
CC -!- SUBUNIT: Homodimer; disulfide-linked.
CC -!- SUBCELLULAR LOCATION: Membrane-associated protein. Golgi.
CC -!- DISEASE: Antigen in chronic rheumatoid arthritis and in the
CC autoimmune disease Sjogren's syndrome.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to licensee@isb-sib.ch).

DR EMBL; X75304; CAAS3052.1; -.
DR EMBL; D25542; BAA05025.1; -.
DR PTR; A56539; A56539.
DR PTR; I52300; I52300.
DR Genew; HGNC:4429; GOLGB1.
DR MIR; 602500; -.
DR GO; GO:0000139; C:Golgi membrane; TAS.
DR GO; GO:0005795; C:Golgi stack; TAS.
DR GO; GO:0016021; C:integral to membrane; TAS.

DR GO:0007030; P:Golgi organization and biogenesis; TAS.

KW Golgi stack; Antigen; Coiled coil; Transmembrane.

FT DOMAIN 1 3235 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 3236 3236 POTENTIAL..

FT DOMAIN 3257 3259 LUMENAL (POTENTIAL).

FT DOMAIN 48 593 COILED COIL (POTENTIAL).

FT DOMAIN 677 1028 COILED COIL (POTENTIAL).

FT DOMAIN 1062 1245 COILED COIL (POTENTIAL).

FT DOMAIN 1301 1779 COILED COIL (POTENTIAL).

FT DOMAIN 1828 3185 COILED COIL (POTENTIAL).

FT DOMAIN 2420 2423 POLY-GLU.

FT DOMAIN 2993 2996 POLY-SER.

FT CONFLICT 1 39 MISSING (IN REF. 3).

FT CONFLICT 215 215 A -> AOLSSM (IN REF. 3).

FT CONFLICT 1765 1765 D -> G (IN REF. 3).

FT CONFLICT 2950 2950 H -> D (IN REF. 3).

SQ SEQUENCE 3259 AA; 376075 MW; 60376A20D8A178DD CRC64;

Query Match

Best Local Similarity 2.1%; Score 190.5; DB 1; Length 3259;

Matches 221; Conservative 207; Mismatches 424; Indels 377; Gaps 49;

QY 169 VNHFEPRAGTIKENSRCPLNEEVIVQARKIFPSVIKYVVMETIWEBEKELPPELQ--IRE 226

Db 12 VLHLSGGDDTDQNMRAPLDPELHQESDMFENNTQEDVQERLAYAE-QLVVELKDIIRQ 70

QY 227 KNERVCYLFNDEHSHSYDHVIVSLQALDCELAEAQLHT-----TADK--EGRRVAKAGA 280

Db 71 KDVL-----QOKDIALOEERKAADNKKLKLHAKAKLTSLNKYEEMKAOGTV 121

QY 281 YAAOCEAKEDIKSHSNVSOHPLHVEVLHSEIMAHOKFALRLGSWWNKINSYSDFRQIF 340

Db 122 LTFEQSEQLSKHDKSSTEEEMIEIKHKLOEKEELISTL----- 163

QY 341 COACREEPDSENPCILSLMLMDAKLYGARKILHELIFSPFMEYMKKLFAMBFVKY 400

Db 164 -QAOL-TQAQAEPQASS-----TEME-----EFVMM 188

QY 401 YKQLOKEYISDDHDSRISITALSVMQMTVPTLARHLIEBON-----VTSVIT 447

Db 189 KOOLQEK-----EEFISTLQALSQTQAQAQVVRKEDARFQTVRLHEDELLQLVT 242

QY 448 ETLLEVLPEYLDNRNKNKFGQYSDKLGKVYAVICDKYLILS---KPTIWERLRM-- 501

Db 243 QADVETEMQOKLURVLORKLEHEESLVGR--AQVDLLOQELTAABQRNQLSQOQQME 300

QY 502 -----QPLEGFRFLKILTCMQGMEIRROVGOHIEVDPDWEAAIAIOMQLKNILLMF 554

Db 301 AEHTNLRNTVTEREESKILLEMELEVAERKLSFH-----NLQEEHHLLLEQF 349

QY 555 QEWCADEEL--LLVAYKECHKAVMRCSTSTFI-SSSKT--VVQSCGHSLETYSRYSEDL 609

Db 350 EQAQAQALESRYSALEQKHKAEMBEKTSHLISLQKTQGLQSQACDALKDQSKLLQDK 409

QY 610 VSIHPLSRTLAGLHVRLSRGA-----VSRL-----HEP---VSFEDFQVEVLVEPLRC 657

Db 410 NQAVQSAQTIQOQEDLOQKQKSEISQFLNRLPLQOQHETASQTSFPDVPVNEGTVTEEN 469

QY 658 LVLVAQVVAEMRRNGLSLISVYFYQDVKCREMYDKDILMLQI-----GAS 705

Db 470 IASLQRRVLENEKEGALLLSI-ELEELKAENEKLSQITLLEAQNRTGEADREYSEIS 528

QY 705 LMD-PNK-----FLLVLQRYELAE 725

Db 529 IVDIANKSSSAEESQDVLENTFSQKHKSLVILLEMKEAQEEIAFLKLOQKRAEEA 588

QY 726 FNKTSTKDDLIKQY--NTLIEMLQVLIYIVGERY----- 760

Db 589 DHEVL---DQEKWKQMEGEGAPIKMKVPLEDTGQDFPLMPNEESSLPAVEKEQASTEHO 645

QY 761 -----VPGVGNVTKEEVMYREIHLHCIEPMPHSA- 790

Db 646 SRTSEEISLNDAGVELKSTQDKGKSLSAVVDIGQCHQDELE-RLKSQLIELELNFHKAQ 704

QY 791 --IAKNLPENENNENETGLENVINKVATFKKPGVSGHGVY-----ELKDESL---KDFNMVYF- 840

Db 705 EIEKNDLDEKAKEISNLNQLIEE---FKKNADNNSSAFTALSEEDQLLSQVKELSMVTE 761

QY 841 --YHYSTQHSKAEHMKKRRKQENK---DEALPPPPPPPEFCPAFSKVINLLNCIDIMYI 895

Db 762 LRAQVKOLEMNLAEARQRRLLDYESQTAHDNLLTE-----QIHSLSTEAKSKDVKEV 814

QY 896 LRTVFE-----POLEGOKDMITWLOMDFVTKRLREKSLCLIVATTSSESITKNDIEITHD 1018

Db 815 LQNELDDVQLQFSEOSTLIRSLQSQONKESEVLEGAERVRHRISSKVEELSQAQSKLE 874

QY 922 ILALG--LLEEK---QLOKAPEE-----EVTDFYIKASRLG---SSAMNIQMLLEK 966

Db 875 ITRKMDQLLEKKRDVETLQOTIEEKDQVQTEISFSMTKVMQVQNEEKFSLGVEIKTLRQ 934

QY 967 LKGI-----POLEGOKDMITWLOMDFVTKRLREKSLCLIVATTSSESITKNDIEITHD 1018

Db 935 LNLSSRAEBAEKQVEEDNEVSSGLKQNYDEM-----SPAGQISKLEQHE 980

QY 1019 -----KEKAERKKRAEAAARLHRQIKMAQMSALQKNFIETHKLMYDNTSEMPGKEDSIM 1073

Db 981 FDLKKENEQKRRKLQAALINRKELLQVRSLEE---ELANLKDESKKEIPLSETERGEV 1037

QY 1074 ESTPAVDYSRIALGPKRGPSVTEKEVLTCLQBE-----EQEVKIENNAMVLS- 1122

Db 1038 EEDKENKEYSE-----KCVTSK-----CQIEIYLVKTISEKEVELQHIRKDLBE 1082

QY 1123 --ACVQKSTALTOHRGKPIELSGEALDPL 1149

Db 1083 KLAEEQFQALVKKOMNOTLQDKTNOIDLL 1111

RESULT 7

SNEI_HUMAN STANDARD; PRT: 8797 AA.

AC Q9NF91; Q94890; Q8NP7; Q8TCPI; Q8WMW6; Q8WMW7; Q8WVF6; Q96N17;

AC Q9C0A7; Q9H525; Q9NS26; Q9NS36; Q9NU06; Q9UJ06; Q9UJ07; Q9ULF8;

DT 15-SEP-2003 (Rel. 42, Created)

DT 15-SEP-2003 (Rel. 42, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Nesprin 1 (Nuclear envelope spectrin repeat protein 1) (Synaptic nuclear envelope protein 1) (Syn-1) (Myocyte nuclear envelope protein 1) (Myne-1) (Enaptin).

GN SYNE1 OR MYNE1 OR KIAA0796 OR KIAA1756 OR KIAA1262.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3), CHARACTERIZATION, AND MUTAGENESIS OF 8758-LEU-CYS-8763.

RC TISSUE=Heart, Placenta, Skeletal muscle, Spleen, and Testis;

RX MEDLINE=21652858; PubMed=11792814;

RA Zhang Q., Skepper J.N., Yang F., Davies J.D., Hegyi L., Roberts R.G., Weissberg P.L., Ellis J.A., Shanahan C.M.;

RT "Nesprins: a novel family of spectrin-repeat-containing proteins that localize to the nuclear membrane in multiple tissues.";

RL J. Cell Sci. 114:4485-4498(2001).

RN [2]

RP SEQUENCE FROM N.A. (ISOFORM 1), AND SUBCELLULAR LOCATION.

RC TISSUE=Heart, Spleen, and Testis;

RX MEDLINE=22296983; PubMed=12408964;

RA Zhang Q., Ragnauth C., Greener M.J., Shanahan C.M., Roberts R.G.;

RT "The nesprins are giant actin-binding proteins, orthologous to Drosophila melanogaster muscle protein MSP-300.";

RL Genomics 80:473-481(2002).

RN [3]

RP SEQUENCE FROM N.A. (ISOFORM 4), AND VARIANT GLY-8323.

RA Braune S., Abraham S., Padmakumar V., Tunngal B., Noegel A.A., Korenbaum E.;

RT "The longest isoform of enaptin/Syne-1, a nuclear envelope associated

protein, binds actin cytoskeleton via the alpha-actinin-like actin-binding domain.";
 Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
 [4]
 SEQUENCE FROM N.A. (ISOFORM 8 AND 9).
 Zhang Q., Shanahan C.M.;
 Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
 [5]
 SEQUENCE FROM N.A.
 Almeida J., Clark S., Griffiths C., Lloyd D., Parker A., Smith M., Tracey A., Williams S.;
 Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 [6]
 SEQUENCE OF 1-856 FROM N.A.
 TISSUE=Kidney;
 Gough L., Fan J., Lisa G., Chu S., Winnick S., Beck K.A.;
 "Golgi localization of syne-1.";
 Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
 [7]
 SEQUENCE OF 28-778 AND 2901-3476 FROM N.A.
 TISSUE=Adrenal gland, and Teratocarcinoma;
 Ninomiya K., Wagatsuma M., Kikkawa E., Omura Y., Abe K., Kamihara K., Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K., Tanemori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Takemoto M., Ota T., Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
 "NEO human cDNA sequencing project.";
 Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 [8]
 SEQUENCE OF 443-8797 FROM N.A. (ISOFORM 5).
 TISSUE=Brain;
 MEDLINE=21082932; PubMed=11214970;
 Nagase T., Kikuno R., Hattori A., Kondo Y., Okumura K., Ohara O.;
 "Prediction of the coding sequences of unidentified human genes. XIX. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro.";
 DNA Res. 7:347-355(2000).
 [9]
 SEQUENCE OF 743-8797 FROM N.A. (ISOFORM 6).
 TISSUE=Brain;
 Ansong W., Wirkner U., Mewes H.-W., Well B., Wiemann S.;
 Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 [10]
 SEQUENCE OF 4219-8797 FROM N.A. (ISOFORM 7).
 TISSUE=Brain;
 MEDLINE=20039619; PubMed=10574462;
 Nagase T., Ishikawa K.-I., Kikuno R., Hirose M., Nomura N., Ohara O.;
 "Prediction of the coding sequences of unidentified human genes. XV. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro.";
 DNA Res. 6:337-345(1999).
 [11]
 SEQUENCE OF 6922-8797 FROM N.A.
 TISSUE=Brain;
 MEDLINE=99087487; PubMed=9872452;
 Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
 "Prediction of the coding sequences of unidentified human genes. XI. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro.";
 DNA Res. 5:277-286(1998).
 [12]
 REVISIONS.
 TISSUE=Brain;
 MEDLINE=22158633; PubMed=12168954;
 Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;
 "Construction of expression-ready cDNA clones for KIAA genes: manual curation of 330 KIAA cDNA clones.";

DNA Res. 9:99-106(2002).
 [13]
 SEQUENCE OF 7631-8797 FROM N.A., AND CHARACTERIZATION.
 MEDLINE=21659781; PubMed=11801724;
 Mislow J.M.K., Kim M.S., Davis D.B., McNally E.M.;
 "Myne-1, a spectrin repeat transmembrane protein of the myocyte inner nuclear membrane, interacts with lamin A/C.";
 J. Cell Sci. 115:61-70(2002).
 [14]
 SEQUENCE OF 8406-8797 FROM N.A.
 Ma F.-R., Zhu L.-P.;
 Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 -!- FUNCTION: Involved in the maintenance of nuclear organization and structural integrity. Probable anchoring protein which tethers the nucleus to the cytoskeleton. Connects nuclei to the cytoskeleton by interacting with the nuclear envelope and with F-actin in the cytoplasm.
 -!- SUBUNIT: Interacts with MUSK, with F-actin via its N-terminal domain, and with LMNA in vitro (By similarity).
 -!- SUBCELLULAR LOCATION: Type IV membrane protein (Potential). The largest part of the protein is cytoplasmic, while its C-terminal part is associated with the nuclear envelope, most probably the outer nuclear membrane. In skeletal and smooth muscles, a significant amount is found in the sarcomeres.
 -!- ALTERNATIVE PRODUCTS:
 Event=Alternative splicing; Named isoforms=9;
 Name=1;
 IsoId=Q8NF91-1; Sequence=Displayed;
 Name=2; Synonyms=Beta;
 IsoId=Q8NF91-2; Sequence=VSP_007130;
 Name=3; Synonyms=Alpha;
 IsoId=Q8NF91-3; Sequence=VSP_007132, VSP_007144;
 Name=4;
 IsoId=Q8NF91-4; Sequence=VSP_007134, VSP_007139, VSP_007140, VSP_007144;
 Name=5;
 IsoId=Q8NF91-5; Sequence=VSP_007135, VSP_007136;
 Note=No experimental confirmation available;
 Name=6;
 IsoId=Q8NF91-6; Sequence=VSP_007137, VSP_007138;
 Note=No experimental confirmation available;
 Name=7;
 IsoId=Q8NF91-7; Sequence=VSP_007141, VSP_007142;
 Note=No experimental confirmation available;
 Name=8; Synonyms=Beta 2;
 IsoId=Q8NF91-8; Sequence=VSP_007131;
 Name=9; Synonyms=Alpha 2;
 IsoId=Q8NF91-9; Sequence=VSP_007133, VSP_007143, VSP_007144;
 -!- TISSUE SPECIFICITY: Widely expressed. Highly expressed in skeletal and smooth muscles, heart, spleen, and peripheral blood leukocytes.
 -!- DOMAIN: The Klarsicht domain, which contains a transmembrane domain, mediates the nuclear envelope targeting.
 -!- SIMILARITY: Belongs to the Nesprin family.
 -!- SIMILARITY: Contains 1 actin-binding domain.
 -!- SIMILARITY: Contains 2 calponin-homology (CH) domains.
 -!- SIMILARITY: Contains 12 HAT repeats.
 -!- SIMILARITY: Contains 1 Klarsicht domain.
 -!- SIMILARITY: Contains 31 spectrin repeats.
 -!- CAUTION: Ref.5 (CAB55865, CAB55866, CAC16280 and CAC16281) sequences differ from that shown due to erroneous gene model prediction.
 -!- CAUTION: Ref.7 (BAB71097) sequence differs from that shown due to a chimeric cDNA.
 -!- CAUTION: Ref.14 sequence differs from that shown due to two frameshifts in positions 8412 and 8784.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

Qy	915	-----MLQAFHILA-----LGLLEK---QQLQKAPBEEVTFDYHKA	950
Db	6033	CEADPBQALQSTLTVLAERMSTIRMKASGKQQLLEKNDQLEQEQALQRYCEA	6092
Qy	951	SRGSSAMNTOMLEKLGKIPQLEGKDMITWILQWFDTVKRLREKSLIVATSSSESI	1010
Db	6093	DEDSWLLSTKATDIALSP-----KEPMDMAQLMDCONMLVEIQKVVALS--ELSV	6145
Qy	1011	KNDEI-----THDKAEAKERRKAAEALRHQKIMAQMSALQKQFIETHKLMYDNTSEMPG	1065
Db	6146	HNENLLEGAHTKDEAEQL-----AGKLR-----LKGSLLELQRALHDKQLNMQG	6192
Qy	1066	-----KEDSIMEEST---PAVSDY---SRIALGPKRGPVSVEKEVLCILQCEQEVKIEIN	1116
Db	6193	TAQEKESVDLDTATQSPGVQEWLAQARTTWTQORSSLOQOKEL-----EQEL-AEQ	6244
Qy	1117	NAMVLSACVQKSTALQTHRCKPTELSEALD-PLFMDPDLAYGTYTTCGSHGVHVCWQK	1175
Db	6245	KSILRSVASGEELIQH---SAAETSGAGEKPDVLQSELGMEGKSSAEDQMR-MKWS	6301
Qy	1176	YFEAVQLSSQORIHVDLFDLESGEYICPLCKSLCNTVPIPIQPQKINSENADALAQLL	1235
Db	6302	LHQ--FESTQKLLQNVEQEQLV---YSRPNLLSGVPLYKGDVPTQDSKAVTSLL	6355
Qy	1236	T-IARNTQTVLARISGVNIRHAKENPIPIFFNOGMGDSITLPHSLISFGVSESIKYSNS	1294
Db	6356	DGLNQAFEEVSSQSG---AKRQS---IHLEQKLYD-----GVSATSTWLDD	6396
Qy	1295	IKEMVILFATTIYRIGLKVPDPRDPVPMLTWSTCAFTQAIENLLG-----DECKPL	1348
Db	6397	VEER--LFVAT-----ALLPEETE-----TCLFNEIILAKDIKEMSEMDKNNKL	6439
Qy	1349	FG-ALQNRQHN-----GLKALMFAVAQRITCQVLIQKHLVLLS	1388
Db	6440	FSQAFFPENGDRVDIEDTLGCLLGRLLDSVYNQR--CHQ--MKERLQQLIN	6488

RESULT 8

CEP2_HUMAN

ID

CEP2_HUMAN

STANDARD;

PRT;

2442 AA.

AC

Q9BV73; O14812; O60588; Q9H450;

DT

15-SEP-2003 (Rel. 42, Created)

DT

15-SEP-2003 (Rel. 42, Last sequence update)

DT

15-SEP-2003 (Rel. 42, Last annotation update)

DE

Centrosomal protein 2 (Centrosomal Nek2-associated protein 1) (C-NAP1)

DE

(Centrosome protein 250) (Centrosome associated protein CEP250).

GN

CEP2 OR CNAP1.

OS

Homo sapiens (Human).

OC

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX

NCBI_TaxID=9606;

OX

[1]

RN

SEQUENCE FROM N.A. (ISOFORM 1), AND AUTOIMMUNE DISEASE.

RP

TISSUE=Cervical carcinoma;

RC

MEDLINE=98165428; PubMed=9506584;

RA

MacK G.J., Rees J., Sandblom O., Balczon R., Fritzler M.J.,

RA

Rattner J.B.;

RT

"Autoantibodies to a group of centrosomal proteins in human autoimmune

RL

sera reactive with the centrosome.";

RL

Arthritis Rheum. 41:551-558(1998).

RL

[2]

RN

SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), INTERACTION WITH NEK2,

RP

AND SUBCELLULAR LOCATION DURING THE CELL CYCLE.

RC

TISSUE=Placenta;

RC

PubMed=9647649;

RA

Fry A.M., Mayor T., Meraldi P., Stierhof Y.-D., Tanaka K., Nigg E.A.;

RA

"C-Nap1, a novel centrosomal coiled-coil protein and candidate

RT

substrate of the cell cycle-regulated protein kinase Nek2.";

RT

J. Cell Biol. 141:1563-1574(1998).

RN

[3]

RN

SEQUENCE FROM N.A.

RP

MEDLINE=21638749; PubMed=11780052;

```

CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AY061755; AAL33798.1; -
DR EMBL; AY061756; AAL33799.1; -
DR EMBL; AF495910; AAN60442.1; -
DR EMBL; AF53142; AAN03486.1; -
DR EMBL; AY184203; AAO27771.1; -
DR EMBL; AY184206; AAO27774.1; -
DR EMBL; AL049548; CAB55865.1; ALT_SEQ.
DR EMBL; AL049548; CAB55866.1; -
DR EMBL; AL078582; CAB87586.1; -
DR EMBL; ALI36079; -; NOT_ANNOTATED_CDS.
DR EMBL; ALI38832; CAC16280.1; ALT_SEQ.
DR EMBL; ALI38832; CAC16281.1; ALT_SEQ.
DR EMBL; ALI357081; -; NOT_ANNOTATED_CDS.
DR EMBL; AL450401; -; NOT_ANNOTATED_CDS.
DR EMBL; AL589963; -; NOT_ANNOTATED_CDS.

Query Match          2.1%; Score 189.5; DB 1; Length 8797;
Best Local Similarity 19.9%; Pred. No. 0.023;
Matches 261; Conservative 217; Mismatches 480; Indels 355; Gaps 68;

QY 254 LDCELAEAQL-----HTTAIDK--EGRRAVAGAY-----AACQAKED 290
Db IDAQLEELQILLTEATNHRQNIEMKAAEQEKYGLYTILPSLSLQAEVALDLKIRDQ 5412

QY 291 IKSHSENVSOHPLHVEVLHSEIMAHOKFALRLGSWNKIMSYSDFRQIFC-QACLRREP 349
Db IDOKIKRIVEQSKATSOELSRQI---OKLADLTILTCLKAKTDNVQAOTDKVLGEEL 5469

QY 350 DSENPCLISRLMLWDAKLYGARKILHELIFSFFMEM--EYKFLFAMEVKKYKLOKE 407
Db IDGNC-----SKLMELDAVQR-----FLEQNGQLGKPLAKKIGK-LTELHQQ 5510

QY 408 YISDDHRSISITALSVQMETVPTLARHLTEEQNVISVITETLEVLPEYLDNRNNKFNQ 467
Db ITIRQAENR-----LSKLNQAASHL-EEYN-----EMLELILKWKAK----- 5547

QY 468 GYSQDKLGRVYAVICDLKYILISKPTIW--TERLRMQF-----LEGFRS 509
Db VLAHGTTAUNASASQLREQYILHQTLLEESKIDSELEAMTE 5588

QY 510 FLAILTCMQGWEIRIQVGQHIIEVDDPWEAAIAIQO-----LKNILLMF 554
Db KQGYLTSVYCTEKMSQOAVELGRETELROMIKIRLQNLQDAADKMKFPAELKQLQAAL 5648

QY 555 QEWCA--CDELLLVAYKE--CHKAVNRCSTSFSSSKTVQSCGHSLETKTSYRVSEDLV 610
Db EQAQYLTSPVGRLSLKEQSLSHRQLLSMESLSLKPQVQVQJCSAL-----RIPEDVV 5703

QY 611 SIHLPSRLTAGLHVRLSRL--CAVSRHL-----EFVSFEDQVEV-----LVEYPLRCLV 659
Db A-SLPLCHAALRLQEEASRLQHTAIQOCNIMQAQVVOYEQYEQEMKHLQOLIGEAHR--E 5760

QY 660 LVAQVVAEMRRNGLSLISQVFFYQVKCREENYDKDIIMQJIGASLMDPN-----KFLLL 715
Db IEDKPAV---TSNIQELQAOISRHEELAOKIKGYEQJIASLSNCKCKMLTKMAKHMTLLT 5817

QY 716 VLQRYELAEAFNKTISTKQDOLIKQYNTLIEML-----QVLIIVGERY--VPQVGN 766
Db VTEVEGLAE-----GTEDLD-----GELLPTPSAHPVMMTAGRCHTLLSPVTE 5862

QY 767 VTKEEYVTRRIIH-LLCIEPMPHSAIKALNPENNETGLNVIN--KVAT----- 814
Db ESKEEGTNEISFPACRSPSP-----VANTDASVNQDIAYQALSERLOTDAAKTHPST 5918

QY 815 -----FKKGVSGHGYVELKD-----ESLKDFNNFYFYHYSKTOHQSAEHMKRRKQEN- 863
Db SASQEEYFEGCLEPSATAKLQDLSRWETLKNV-----ISEKORTLYEALERQOKYQDSL 5972

QY 864 -----KDEA-----LPPPPPPFPCPAPSKVINLNCDIMMYILTVTFERAIDTDSNLWTEG 914
Db QSTISTKWEAIELKLSPPGSPGSPQMAHQALMDEILMLQDEINELLOSSLAEEIVSES 6032

```


RA Dufosse-Laurent V., Ferron C., Lechaplais C., Louesse C., Muelet D.,
RA Magdelenat G., Pateau E., Petit E., Sirvain-Trukniewicz P., Trybou A.,
RA Vega-Czarny N., Bataille E., Bluet E., Bordelais I., Dubois M.,
RA Dumont C., Guerin T., Haffray S., Hammadi R., Muanga J., Pellouin V.,
RA Robert D., Wunderle E., Gauguet G., Roy A., Sainte-Marthe L.,
RA Verdier J., Verdier-Discala C., Hillier L., Fulton L., McPherson J.,
RA Matsuda F., Wilson K., Scarpelli C., Gyapay G., Fulten L., McPherson J.,
RA Quetier F., Waterston R., Hood L., Weissbach J.,
RA "The DNA sequence and analysis of human chromosome 14.";
RA Nature 421:601-607(2003).
RA [6]
RA SEQUENCE FROM N.A. (ISOFORM 8).
RA TISSUE=Brain;
RA MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Heiton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzyzanski M.I., Skalska U., Smalhus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length human
RA and mouse cDNA sequences.";
RA Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RA [7]
RA SEQUENCE FROM N.A. (ISOFORM 3), AND SEQUENCE OF 1-956 AND 5133-6885
RA FROM N.A.
RA TISSUE=Spleen, and Tongue;
RA Jikuya H., Takano J., Nomura N., Kikuno R., Nagase T., Ohara O.,
RA Ninomiya K., Wadatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
RA Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,
RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RA "NEDO human cDNA sequencing project.";
RA Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
RA [8]
RA SEQUENCE OF 5709-6885 FROM N.A. (ISOFORM 2).
RA TISSUE=Brain;
RA MEDLINE=99246063; PubMed=10231032;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirose M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RA "Prediction of the coding sequences of unidentified human genes. XIII.
RA The complete sequences of 100 new cDNA clones from brain which code
RA for large proteins in vitro.";
RA DNA Res. 6:63-70(1999).
RA [9]
RA REVISTONS.
RA TISSUE=Brain;
RA MEDLINE=22158633; PubMed=12168954;
RA Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;
RA "Construction of expression-ready cDNA clones for KIAA genes: manual
RA curation of 330 KIAA cDNA clones.";
RA DNA Res. 9:99-106(2002).
RA [10]
RA SEQUENCE OF 5754-6885 FROM N.A.
RA MEDLINE=21154917; PubMed=11230166;
RA Wienann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,
RA Ansgore W., Boecker M., Bloecher H., Bauersachs S., Blum H.,
RA Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,

RA Meves H.-W., Ottenwaelder B., Obermaier B., Tampe J., Heubner D.,
RA Wambutt R., Korn B., Klein M., Poustka A.;
RA "Towards a catalog of human genes and proteins: sequencing and
RA analysis of 500 novel complete protein coding human cDNAs.";
RA Genome Res. 11:422-435(2001).
RA CC -!- FUNCTION: Involved in the maintenance of nuclear organization and
RA structural integrity. Probable anchoring protein which tethers the
RA nucleus to the cytoskeleton. Connects nuclei to the cytoskeleton
RA by interacting with the nuclear envelope and with F-actin in the
RA cytoplasm.
RA CC -!- SUBUNIT: Interacts with F-actin via its N-terminal domain.
RA CC -!- SUBCELLULAR LOCATION: Type IV membrane protein (Potential). The
RA largest part of the protein is cytoplasmic, while its C-terminal
RA part is associated with the nuclear envelope, most probably the
RA outer nuclear membrane. Remains associated with the nuclear
RA envelope during its breakdown in mitotic cells.
RA CC -!- ALTERNATIVE PRODUCTS:
RA Event-Alternative splicing; Named isoforms=9;
RA Name=1;
RA IsoId=Q8WXH0-1; Sequence=Displayed;
RA Name=2;
RA IsoId=Q8WXH0-2; Sequence=VSP_007164, VSP_007166;
RA Note=No experimental confirmation available;
RA Name=3;
RA IsoId=Q8WXH0-3; Sequence=VSP_007155;
RA Note=Produced by exon skipping that results in a frameshift. No
RA experimental confirmation available;
RA Name=4; Synonyms=Beta;
RA IsoId=Q8WXH0-4; Sequence=VSP_007156;
RA Name=5; Synonyms=Alpha;
RA IsoId=Q8WXH0-5; Sequence=VSP_007157, VSP_007164, VSP_007165;
RA Name=6;
RA IsoId=Q8WXH0-6; Sequence=VSP_007158, VSP_007165, VSP_007166;
RA Note=No experimental confirmation available;
RA Name=7; Synonyms=Gamma;
RA IsoId=Q8WXH0-7; Sequence=VSP_007154, VSP_007163;
RA Name=8;
RA IsoId=Q8WXH0-8; Sequence=VSP_007161, VSP_007162;
RA Note=No experimental confirmation available;
RA Name=9; Synonyms=NUANCE-N-33;
RA IsoId=Q8WXH0-9; Sequence=VSP_007159, VSP_007160;
RA CC -!- TISSUE SPECIFICITY: Widely expressed, with higher level in kidney,
RA adult and fetal liver, stomach and placenta. Weakly expressed in
RA skeletal muscle and brain. Isoform 5 is highly expressed in
RA pancreas, skeletal muscle and heart.
RA CC -!- DOMAIN: The Klarsicht domain mediates the nuclear envelope
RA targeting
RA CC -!- SIMILARITY: Belongs to the Nesprin family.
RA CC -!- SIMILARITY: Contains 1 actin-binding domain.
RA CC -!- SIMILARITY: Contains 2 calponin-homology (CH) domains.
RA CC -!- SIMILARITY: Contains 1 Klarsicht domain.
RA CC -!- SIMILARITY: Contains 13 leucine-rich (LRR) repeats.
RA CC -!- SIMILARITY: Contains 9 spectrin repeats.
RA -----
RA This SWISS-PROT entry is copyright. It is produced through a collaboration
RA between the Swiss Institute of Bioinformatics and the EMBL outstation -
RA the European Bioinformatics Institute. There are no restrictions on its
RA use by non-profit institutions as long as its content is in no way
RA modified and this statement is not removed. Usage by and for commercial
RA entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
RA or send an email to license@isb-sib.ch).
RA -----
RA EMBL; AF435010; AAL33547.1; -;
RA EMBL; AF435011; AAL33548.1; -;
RA EMBL; AY061757; AAL33800.1; -;
RA EMBL; AY061758; AAL33801.1; -;
RA EMBL; AY061759; AAL33802.1; -;
RA EMBL; AF495911; AAN60443.1; -;
RA EMBL; AL117404; CAB55905.1; -;
RA EMBL; AL162832; -; NOT_ANNOTATED_CDS.
RA EMBL; AL355094; -; NOT_ANNOTATED_CDS.
RA -----

Query Match 1.9% Score 177; DB 1; Length 6885;

Best Local Similarity 18.2%; Pred. No. 0.094;

[illegible]

Db	1837	SKLLNDQCKNFNDWFSIKVNLKFCFESSETKKSEVQKLOKLSDFLFLGRNSKIKQVDS	1839
Qy	1120	VLSACVQKSTALTQHRGKPIELSGEALDPLFMDPLAYGYTSGCHVMHVCWQKFEA	1179
Db	1897	VLK-----HVKKHLPK-----AHVKELISW-----	1916
Qy	1180	VQLSSQORIHVDLFLSEGEVLCPLCKSLCNTVPIPILOPKINSENADALAQL	1235
Db	1917	--LVGQE-----FELEKMSIC-----QARAKELEDSLOQLRLQD	1950
Qy	1236	---TLARIQTVLARIISYINIRHAKGNPIPIFNQMGDSTLFEHSILSGVSSSTK-Y	1291
Db	1951	DHRNLRKWLNTQBEKWKGMPEPGKTE-----LFCQALARKREQFESVAQ--LNNSLKEY	2003
Qy	1292	SNSIKEMVILLFATTI---YRIGLK--VPPDERDPRV-----MLTWSTCAFTI	1334
Db	2004	GFTEEEIIMEATCLMDRYOTLLKQLSEIEEDKLLPTEOSFNDLAHDVIHW-----I	2057
Qy	1335	QAIENLL-----GDEGK-PLFGALQNROHNGLKALMQFAVAQRITCPQVLIQKHLVRLLSV	1389
Db	2058	KEIKESLVLNLSSEGKMPLEIRIQ-----IKEI-----ILLKPSGDARIET	2099
Qy	1390	VLPNIKESDTPCLLSIDLFIHVLGCVAFPSLYWDPDVLOPSVSSSYNHLXLFHLITM	1449
Db	2100	IMKQAESAP-----LVQKLTDISNQWNTLHL-----ASTY-----L	2134
Qy	1450	AHMLQIILLTVDTGLPLAQVQED	1471
Db	2135	SHOEKLLL---EGEKYLOSKED	2153

RESULT 10

MY5A_CHICK

ID	MY5A_CHICK	STANDARD;	PRT;	1829 AA.
AC	Q02440;			
DT	01-JUN-1994 (Rel. 29, Created)			
DT	01-JUN-1994 (Rel. 29, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Myosin Va (Myosin 5A) (Dilute myosin heavy chain, non-muscle) (Myosin heavy chain P190) (Myosin-V).			
DE	heavy chain P190) (Myosin-V).			
DE	heavy chain P190) (Myosin-V).			
DE	heavy chain P190) (Myosin-V).			
OS	Gallus gallus (Chicken).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;			
OC	Gallus.			
OX	NCBI_TaxID=9031;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Brain;			
RX	MEDLINE=93012002; PubMed=1383040;			
RA	Sanders G., Lichte B., Meyer H.E., Killmann M.W.;			
RA	Espreafico E.M., Cheney R.E., Matteoli M., Nascimento A.A.,			
RA	de Camilli P.V., Larson R.E., Mooseker M.S.;			
RT	"Primary structure and cellular localization of chicken brain			
RT	myosin-V (p190), an unconventional myosin with calmodulin light			
RT	chains."			
RL	J. Cell Biol. 119:1541-1557(1992).			
RN	FEB5 Lett. 311:295-298(1992).			
RN	[2]			
RC	SEQUENCE FROM N.A., AND SEQUENCE OF 1154-1163.			
PC	TISSUE=Brain;			
RX	MEDLINE=93107155; PubMed=1469047;			
RA	Espreafico E.M., Cheney R.E., Matteoli M., Nascimento A.A.,			
RA	de Camilli P.V., Larson R.E., Mooseker M.S.;			
RT	"Primary structure and cellular localization of chicken brain			
RT	myosin-V (p190), an unconventional myosin with calmodulin light			
RT	chains."			
RL	J. Cell Biol. 119:1541-1557(1992).			
CC	-!- FUNCTION: PROCESSIVE ACTIN-BASED MOTOR THAT CAN MOVE IN LARGE			
CC	STEPS APPROXIMATING THE 36-NM PSEUDO-REPEAT OF THE ACTIN FILAMENT.			
CC	MAY BE INVOLVED IN MELANOSOME TRANSPORT, OR ALTERNATIVELY, IT MAY			
CC	BE REQUIRED FOR SOME POLARIZATION PROCESS INVOLVED IN DENDRITE			
CC	FORMATION (BY SIMILARITY).			
CC	-!- SUBUNIT: MAY BE A HOMODIMER, WHICH ASSOCIATES WITH MULTIPLE			
CC	CALMODULIN OR MYOSIN LIGHT CHAINS.			
CC	-!- SUBCELLULAR LOCATION: GOLGI-DERIVED CYTOPLASMIC MEMBRANES			

CC CC (POTENTIAL).

CC CC -!- TISSUE SPECIFICITY: NEURONAL AND NONNEURONAL CELLS OF THE BRAIN.

CC CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.

CC CC -!- SIMILARITY: Contains 1 dilute domain.

CC CC -!- SIMILARITY: Contains 6 IQ domains.

CC CC -----

CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC CC the European Bioinformatics Institute. There are no restrictions on its

CC CC use by non-profit institutions as long as its content is in no way

CC CC modified and this statement is not removed. Usage by and for commercial

CC CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC CC or send an email to license@isb-sib.ch).

CC CC -----

DR EMBL; X67251; CAA47673.1; -

DR EMBL; Z11718; CAA77782.1; -

DR PIR; S19188; S19188.

DR HSSP; P10587; 1BR2.

DR InterPro; IPR002710; DIL.

DR InterPro; IPR000048; IQ-region.

DR InterPro; IPR001609; myosin_head.

DR Pfam; PF01843; DIL; 1.

DR Pfam; PF00612; IQ; 6.

DR Pfam; PF00063; myosin_head; 1.

DR PRINTS; PR00193; MYOSINHEAVY.

DR ProDom; PD003376; DIL; 1.

DR ProDom; PD000355; myosin_head; 1.

DR SMART; SM00015; IQ; 6.

DR SMART; SM00242; MYSC; 1.

DR PROSITE; PS50096; IQ; 6.

DR Myosin; Repeat; ATP-binding; Calmodulin-binding; Actin-binding;

CC CC Coiled coil; Phosphorylation.

CC CC MYOSIN HEAD-LIKE.

FT DOMAIN 1 766

FT DOMAIN 767 789

FT DOMAIN 790 814

FT DOMAIN 815 837

FT DOMAIN 838 862

FT DOMAIN 863 887

FT DOMAIN 888 915

FT DOMAIN 916 1239

FT DOMAIN 1315 1419

FT DOMAIN 1661 1766

FT NP_BIND 163 170

FT DOMAIN 644 666

FT MOD_RES 1734

FT CONFLICT 1142 1142

CC CC E -> EQ (IN REF. 2).

CC CC SEQUENCE 1829 AA; 212381 MW; 0538B278DFC09F6E CRC64;

Query Match 1.9%; Score 176; DB 1; Length 1829;

Best Local Similarity 18.0%; Pred. NO. 0.015;

Matches 283; Conservative 248; Mismatches 578; Indels 464; Gaps 76;

Qy 99 GAFOLCGRVFKSGTYSYCRD---CAI---DPTCVLC---MDCFDQSVHKNHRYKMH 147

Db 324 GIFRLAGILHLGNVFAASRDSDCAIPPKHDLPTIFCDLMGVD-YEEMAWLCHKRLAT 382

Qy 148 STGGGFCDCGTEAWKTPFCVFNHEPGRAGTIKNSRCPLNEEVYQARKIPPSVIK--- 204

Db 383 A-----TETY-IKPISKHAINARDALAKHIYANFNWIVDHVNKALHSTVKQHS 431

Qy 205 YVEMTWEEBKELPPELQIREKERYCYVFNDE--HHSYDHYIYSQRLDCLAEQA 262

Db 432 FIGVLDIYGF-----TPEINSFEQ--FCINYANEKLOQFNMHVFKLEQE---EYMREQ 481

Qy 263 LHTTAIDKEGR-----AVKAGAYAAOEAKEKIDKSHSNVSOHPLHVEVLHSEIMAHQ 316

Db 482 IPWTLDIFYDNPICINLEAKMGVLDLDECKMPKGSDDTWAQKLYNTHL----- 533

Qy 317 KFAL-----RLGSWMNK--IMSYSSDFRQIFQACLRPEEDSENPCILSRMLWDKLYKG 370

Db 534 KCALFEKPRLS--NKAFIHKHFKADKVEYQCEGFLEKNKD-----VYEE 575

Qy 371 ARKILHELIFSFTMEMEYKKLFAMEFVYKYLQLOKEYSIDDDHRSISITALS-----VQM 426

Db 576 QIKVL-----KSSKKFKLLPE--LFODEKAISPTISATPSGRVPL 613

Qy 427 FTVF-----TLARHL-----IEQNVISVITETLLLEVPEYLD--RNNKFNFGYS 470

Db 614 SRTVPKAKARPGQTSKEHKKTGVGHFRNSLHLLMETLNATTPHYVRCIKPNDPKFPPTF 673

Qy 471 QDKLG-----RVYAVICDLAYILISKPTIATERLRFMOFLGFRSFLK---IL-----TCMQ 518

Db 674 DEKRAVOQLRACGVLETIRISAAGFSPRWYQ---EFFSKRYVLMKOKDVLSDRKOTCKN 730

Qy 519 GMEEI-----RRQVG-----OHIEVDPDWEAAIAIOMQLKNILLMFOWCA 559

Db 731 VLEKLILDKDKYQFGKTKIFFRAGQVAYLEKTRADKLRAACIRIQKTIRGWLRRKMYRM 790

Qy 560 CDEELLVAYKECHKAMVRCSTSFSSKT--VVQSCGHS--LFTKSYRVSDELVSHLPL 616

Db 791 RRAAITIQRVVRGHOA--RCYATFLRTRAAIIIQKFORMYVVRKRYQCMRDAT---IAL 845

Qy 617 SRTLAGLHVRLSRLGAVSRLEHFVSFEDFQVEVLVEPLCLVLVAQVVAEMRRRGLSL 676

Db 846 QALLRGYLVN--NKYOMMLREKSI-----IIQKHVRG----- 877

Qy 677 ISOVFYQDVKC-----REEMYDKDIIMLIQIGASLMDPNKFLLLVLQRYELAEAFNK 728

Db 878 LARVHYHRTLKAIVYLQCCYRRMMARELKLKLEARSERYKKLHGLE-----NK 929

Qy 729 --TISTKQDOLIKOYNTLIEEM--LOVLIYIGERYVPGV-----GNVTKEEVTM 774

Db 930 IMQLOKIDEONKEYKSLEKMNLTITYSTETEKLRSDVERLRMSEEAKNATNRVLSL 989

Qy 775 REIHLLECPMPHSAIAKNLPENENNETGLENVINKVATFKPGVSGHGVYELKDESL- 833

Db 990 QEEI-----AKRLKELHQTOQTEKKTIEMADKYHETELQVLS-----ELKQNTL 1034

Qy 834 -----KDFNMVYHYSKTQHSKAEHMOKK---RRKO---ENKDEALPPPPPPFCFAFSK 882

Db 1035 LKTEKEELNRRIHQAK---EITETWEKKLVEETKOLELDLNDERL-----R 1078

Qy 883 VINLLN-----CDIMMYILRTVPERAIDTDSNLWTEGMLQMAFHILA--- 924

Db 1079 YQNLLNEFSRLEERYDDLKDENNLMVSIKPGHKRTDSTHSS--NESEYTFSSSEITAEAD 1136

Qy 925 IGLLEEKQOLQKAPEEVTFDFYHKASRLGSSAMNOMLEKLGIPQLEGQDMTWIL 984

Db 1137 LPLRMEEPSEKAP-----DMSLSFLKQKRYTELEQEKOSLQ--- 1174

Qy 985 QMFDTVKLRKESCLIVATTSGSEIKNDEITHDKEKAERKAEARLHROKIMQMSA 1044

Db 1175 ---DELDRKEQALRAKAKEERPPIRGAEEYESLK---ROELES---ENKKLKHELNE 1225

Qy 1045 LQKNFIETHKLMYDNTSEMPGKEDSIWEEESTPAVDYSRIALGPKRGPVTEKEVITCI 1104

Db 1226 LQKALTETR-----APEVTAQAPAY-RVLL-----DQLTSV 1256

Qy 1105 LCOEEOEVKIENNAWVLSACVOKSTALTQHRGKPIELSGALDPLPMDPLAYGTYGSC 1164

Db 1257 --SEELEVRKEEVLILRSOLVSOKAEI-----QPKEDKNTMTDSTILLEDVQRMKQGEI 1309

Qy 1165 GHVMAVCQKYFEAVQLSSQORIHYDLPDESSEYLCPLCKSL-----CNTVPIIP 1217

Db 1310 AQAYIGLKTNRLLLESQLOSKSHENELESRGE-----IQSLKEENNQQOQLLAQNQL 1364

Qy 1218 LQPO-----KINSNADALAQL-----LTARMTQTVLARIISGYNIRHAKG 1258

Db 1365 LPEAREIASLOHEITRLTNENLDLMEQLEKQDKTVRKLLKQKLVFAKKKILEVCOMEN 1424

Qy 1259 ENPIPIFFNOGMDSTL-----EFHSILSFGVESSIK----- 1290

Db 1425 ISPGQII-----DEPIRPVNIPRKEKDFQGMLEYKKEDEQKLVKNLLELKPGRVAVNL 1478

Qy 1291 -----YNSNIKEMVILFATTIYRIGLVPPDERPRVPMLTW--STCA 1331

```
Db 1479 IPGLPAYILPMCVRHADYLNDDQVRSLTSTINGI-KKVLKRRGDDFETVSWLSNTR 1537
QY 1332 FTIQAENLGLDGEKPLFGALQNRHNGKALMQFAVA---QRITCPQVLIQKHLVRLLS 1388
Db 1538 F-LHCQKQSGEGFKWKNPNRNEH-----CLTNFDLAEYRVLSDLAIQIYQQLVRVLE 1592
QY 1389 VVLPNIKSEDTPLLSDIFLHVILVGAVALPPLSYWDPDVL--QPSSVS-----SSY 1438
Db 1593 NIL-----QPMIVSGMLEHETIQVSGV-----KPTGLRKRTSSIADEGYTLDSII 1639
QY 1439 NHLYLFLHLMTHAH 1451
Db 1640 ROLNSPHSVMCQH 1652

RESULT 11
REST_HUMAN
AC P30622;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Restin (Cytoplasmic linker protein-170 alpha-2) (CLIP-170) (Reed-
DE Sternberg intermediate filament associated protein).
GN RSN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Peripheral blood monocytes;
RX MEDLINE=92289675; PubMed=1600942;
RA Bilbe G., Delabie J., Bruegggen J., Richener H., Asselbergs F.A.M.,
RA Cerletti N., Sorg C., Odink K., Tarcsay L., Wiesendanger W.,
RA de Wolf-peeters C., Shipman R.;
RT "Restin: a novel intermediate filament-associated protein highly
RT expressed in the Reed-Sternberg cells of Hodgkin's disease.";
RL EMBO J. 11:2103-2113(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92405160; PubMed=1356075;
RA Pierre P., Scheel J., Rickard J.E., Kreis T.E.;
RT "CLIP-170 links endocytic vesicles to microtubules.";
RL Cell 70:1887-900(1992).
CC -!- FUNCTION: SEEMS TO BE A INTERMEDIATE FILAMENT ASSOCIATED PROTEIN
CC THAT LINKS ENDOCYTIC VESICLES TO MICROTUBULES.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC, ASSOCIATED WITH THE
CC CYTOSKELETON.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=P30622-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=P30622-2; Sequence=VSP_000765;
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE REED-STERNBERG CELLS
CC OF HODGKIN'S DISEASE.
CC -!- SIMILARITY: Contains 2 CAP-Gly domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.
CC -----
DR EMBL; X64838; CAA46050.1; -
DR EMBL; M97501; AAA35693.1; -
DR PIR; S22695; S22695.
DR Genew; HGNC:10461; RSN.
DR MIM; 179838; -
DR GO; GO:0005768; C:endosome; TAS.
```

```
DR GO; GO:0005882; C:intermediate filament; TAS.
DR GO; GO:0015630; C:microtubule cytoskeleton; TAS.
DR GO; GO:0008017; F:microtubule binding activity; TAS.
DR GO; GO:0006899; P:non-selective vesicle transport; TAS.
DR InterPro; IPR000938; CAP-Gly.
DR InterPro; IPR001878; Znf.CCHC.
DR Pfam; PF01302; CAP_GLY_2.
DR SMART; SM00343; Znf.C2HC; 1.
DR PROSITE; PS00845; CAP_GLY_1; 2.
DR PROSITE; PS00845; CAP_GLY_2; 2.
KW Cytoskeleton; Microtubules; Coiled coil; Repeat; Alternative splicing.
FT DOMAIN 78 120
FT DOMAIN 143 204
FT DOMAIN 232 274
FT DOMAIN 304 331
FT DOMAIN 350 1342
FT DOMAIN 1408 1421
FT VARSPLIC 457 491
FT FTId=VSP_000765.
FT CONFLICT 1069 1069 D -> E (IN REF. 2).
FT SEQUENCE 1427 AA; 160989 MW; 0A4F16DD94254E8 CRC64;
Query Match 1.9%; Score 175.5; DB 1; Length 1427;
Best Local Similarity 19.4%; Pred No. 0.011;
Matches 200; Conservative 160; Mismatches 344; Indels 329; Gaps 45;
QY 172 EPGRAGTIKENSRCPLNEEVIVQARKIFPSVIKYVEMTIWEEKELPPELQ-----IREK 227
Db 488 EDTRVATVSEKSR-----IMELNDLALRVQVEAELRRR 521
QY 228 NERYCVLFNDEHSHYDHYVYLSQALDC-----ELAAEQLHTTAIDKEGRRVAKGA 280
Db 522 LESNKPAGDVDMSSLSLQELSSLOKLEVTTRDHOEITSLKEHFGARETHQKEIK-L 580
QY 281 YAAQCQ---EAKEDIKS-----HSENVQHPHLHVEHSEIMAHQKFAALRG-SWNNKIMS 331
Db 581 YTATEKLSKENESLSKLEHANKENSVDIALWKSLETAIAHQOAMELKVSFSGKLG 640
QY 332 YSSDFRQIFCQ-----ACLRPEPSENPCILSRMLMDAKLYKGARKILHELI 379
Db 641 ETAEFAELKTQIEKMLDYQHEIENLQNOQDSR----- 674
QY 380 FSSFFMEMEYKLFAMEVKYKQLQKEYISDDHRSISITALSVMFTVPTLARHLIEE 439
Db 675 -AAHAKEMEALRAKMLKVIK-EKENSLEAIRSKLDAED-----OHLVEM 717
QY 440 QNVLSVITET-----LLEVLPEYLDNNKE--NFOGY---SODKLRVYVAVICDLKYILI 489
Db 718 EDTLKLQEAIEIKVKEVLEQLAKCNEQTKVIDNFTSQLKATEKL-----LDL----- 765
QY 490 SKPTIWTIRLMQFLEGRSFLKILTCMQGMEIRROVQGHIEVDPDWEA--AIATQMOL 547
Db 766 -----DALRKASSEG-KSEKKL--RQOLEAAEKQI-KHLEIKNAESSKASITREL 814
QY 548 KNILLMPQWCACDEELLVAYKECHKAVMRCSTSFSSKTVVQSCGHSKTSYRVSE 607
Db 815 QG-----RELKLTNLQE-----NLSEVSQVETLEKQLILKEKFAEASE 854
QY 608 DLVSIHPLSRTLAGLHVRLSRGLGAVSRHLFEVDFEQVEVLVEYPLRCLVLVAQVVAE 667
Db 855 EAVSVQSRSMQETVKNLKHQKEQFNMLSS-----DLEKL----- 887
QY 668 MWRRNGLSLSQVYYQD-----VKCREEMVDKI--IMLQIGASLMDPNKFLLLVLOR 719
Db 888 ---RENIADMEAKPREKDEREEQLIKAKEKI-ENDIAEIMKMSGDNSQLTK----- 935
QY 720 YELAAEFNKTSTKDQDLIKOYNTLIEMLQVLIYIVGVRVPGVGNVTKEVMTWEITH 779
Db 936 -----MNDELRLKERDV-----EELQIKLTKANEN--ASFLOKSIEDMTVK----- 974
QY 780 LLCIEPMPHSAIAKNLPENENNETGLENVINKVATFKKPGVSGHGVYELKDSLKDFNMY 839
Db 975 ---AEQSQQBAAKKHEEEKLEKRLSDLEKKMET-----SHNQOQLKAR----- 1017
```


QY 840 FYHYSKTOHSAEAMQKRRQENKDEALPPPPPEPCAFSKVINLLNCDDIMYILRTV 899
DB 1018 --YERATSETKTHEE-----ILQNL 1036
QY 900 PERAIDTSDNLTWETGMLQMAFHIALGLLBEKQOLQKAPBEVTFDFYHKASRLGSSAMN 959
DB 1037 QKTLLDTEKLL--KGAREE-----NSGLLQLEELRKQADKA-----KAAQTAEDAMQ 1082
QY 960 I--OMLLEKLGIPQLGSGQDMITWILQMPDVTVKRLKESCLIVATTSGESIKNDEITH 1017
DB 1083 IMEQTKETKTETASLEDTKOTNAKLQNELDTLKENLKNVEELNKKSKELLTVENQKKEE 1142
QY 1018 DKEAKRKRKAARLRHQKIMAQMSALQKNFIETHKLMYDNTSEMPGKEDSI-----ME 1072
DB 1143 FRKEIETLKQAAQK-----SQLSALQE--ENVKL-----AEELGRSDREVTSQKLE 1189
QY 1073 EESTPAVDSYRIALGPKRGPSTVEKEVFCILCOEBOEVKIENMAMVLSACV--OKSTAL 1131
DB 1190 EER-----SVLNNQLLEKKRKESKFIKA-----DDEKASLOKSISITSALLTEKDAEL 1238
QY 1132 TOHREKPIELSGE 1144
DB 1239 EKLRENEVTURGE 1251

RESULT 12

MY52_SCHPO
ID MY52_SCHPO STANDARD; PRT; 1526 AA.
AC Q9USIG; P78969;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Myosin type II heavy chain I.
GN MY02 OR SPCC645.05C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A., AND FUNCTION.
RC STRAIN=972;
RX MEDLINE=98075862; PubMed=9415380;
RA May K.M., Watts F.Z., Jones N., Hyams J.S.;
RT "type II myosin involved in cytokinesis in the fission yeast,
Schizosaccharomyces pombe.";
RL Cell Motil. Cytoskeleton 38:385-396(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds R., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Borzym K., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,

RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -!- FUNCTION: REQUIRED FOR CELL DIVISION. IT IS A COMPONENT OF THE
CC CDC12 'SPOT', A STRUCTURE THOUGHT TO MARK THE SITE OF SEPTATION.
CC MAY WORK IN CONJUNCTION WITH MYO3.
CC -!- SUBUNIT: BINDS TO CDC4 AND RLCL.
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -!- SIMILARITY: Contains 1 IQ domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U75357; AAC49908.1; -;
DR EMBL; AL049498; CAB39901.1; -;
DR PIR; T41522; T41522.
DR HSP; P08799; IMND.
DR GeneDB_SPombe; SPCC645.05c; -;
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR004009; Myosin_N.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS0096; IQ; 1.
KW MYOSIN; Actin-binding; ATP-binding; Coiled coil; Calmodulin-binding;
KW Alkylation.
FT DOMAIN 1 757 MYOSIN HEAD-LIKE.
FT DOMAIN 758 787 IQ.
FT DOMAIN 788 875 COILED COIL (POTENTIAL).
FT NP_BIND 170 177 ATP (POTENTIAL).
FT DOMAIN 634 656 ACTIN-BINDING (BY SIMILARITY).
FT DOMAIN 734 748 ACTIN-BINDING (BY SIMILARITY).
FT MOD_RES 674 674 ALKYLATION (BY SIMILARITY).
FT CONFLICT 1337 1337 S -> R (IN REF. 1).
SQ SEQUENCE 1526 AA; 176430 MW; 176430 MW; D71D51D6578192BA CRC64;
Query Match 1.9%; Score 175.5; DB 1; Length 1526;
Best Local Similarity 19.9%; Pred. No. 0.013;
Matches 230; Conservative 147; Mismatches 425; Indels 351; Gaps 54;
QY 44 FUHLAQLVPEITYFAEMDPLEKQESVOMSIPTLEWLYFGED--PDI-CLEKLKHSGA 100
DB 473 FNHHFVLRQEEYM-----KEEIV-----WDFIDFGHDLOPTIDLEKANPIGI 516
QY 101 FOLCGRVFSGGTYTSCRD-C---AIDPTCVLCHDCQDSVHKHRYKMHSTGGGF- 153
DB 517 L-----SCLDEECVMPKATDATFTSKLDALWR---NKLKYPKPFKADQGI 560
QY 154 -----CDCG-DTEAMKGTFCVNHPEGRAGTIKENSRCPLNEEIVQARIFPSPVIKYV 207
DB 561 LFHYAADVPYSTEGL-----LEKNTDPLNEV---AKLLAOSTKNKHA 600
QY 208 EM-----TWEEKEKLPPEL-----QIREKNERYV-CVLFNDEH-- 240
DB 601 TLFSDYQETETKTVRCRTKGLFRTVAQRHKEQLNQLMNFNSTQPHFRCIYVNEEKM 660
QY 241 HSYDHYVLSQALDCELAQALHTTATDKEGRAVKAAYACQAEAKEDIKSHSENVSQ 300
DB 661 HTFNR-----PLVLGQLRCNGV-LEGIRITRAGF-----PNR 691
QY 301 HPLHVEVLHSEIMAHQKFAIRLGLSWMNKIM-----SYSSDFRQIFCOACLRPEPD 350

```
Db 692 LPNDFRVEYIMAHLPDPTGYVESRRASVMLEELKIDEASYRIGVSKIFFKAGVLALE 751
QY 351 SENPCLISRLMWDKALYKARKILHELIFSSFMEME-----YKKLFAMEFVK 399
Db 752 ERRVATLQRLM---TMLQTRIRGFLQKIFQKRLKDIAIKLQANLVYNEERTFPWAK 808
QY 400 YKQKQKEYSIDHDHRSISIT-----ALSVQMTVPTTLARHLIEEONVSVITETL 450
Db 809 LFNRLPLSSQNDKQLAKRDAEIIELKQONSKSEVEDRLVETNNSUTAV-ENL 867
QY 451 L-----EVLPEYLDLDR-----NNKFNFGYSQDKLGRVAVICDLKYIILSKPTIW 495
Db 868 LITERAIALDKEILLRRTORLANIEDSFSETKQONENLQRESA---SLKQI----- 916
QY 496 TERLRMQFLEGRFSFKILITCMQGMEEIRROVQGH-----IEVDPDWEAAIAIQMLKNIL 551
Db 917 NNELESELLE---KTSKVETLLSEQNELKESLLEKDLDDTKGELESRENNATVLSEK 973
QY 552 LMFOEWCACDEBLLLVAYKECHKAVMRCSTSPISSSKTVVQSCGHLETKSYRVSEDLVS 611
Db 974 AEFNFOCKSLQETIYTKDAELDKL-----TKYISDYKTEIQ-----EMRLTNQKMKESIQ 1024
QY 612 IHLPLSRTLGLHLVRLSLGAV-SRLHEFVSPFEDEOVELVPEYPLCLVLVAOVVAEMWR 670
Db 1025 QEGSLSLSLK---RVKKLERENSTLISDVSLKQKEEL-----SVLKGVELTI 1071
QY 671 RNGLSLISQVFIYQ-DVKQ-----REMYDKDIIMLQIGASLMDPNKFLLLVLQRYEL 722
Db 1072 NN---LEEKVNYLEADVQKPLKKELESINDKD-----QILYQL 1107
QY 723 AAFNKTISTKQODLIKQYNTLIBEMQLVLIIVGERYVPGVGNVTKVEYTMREIHLIC 782
Db 1108 QATKNKELEAKYKECLNNIKSTKELEN-----KEEQONLSDSALKYIELQEI----- 1156
QY 783 IEPMPHSAIAKNLPENENNETGLENVINKVATFKPGVSGHGVYELKDESLKDFNMYFYH 842
Db 1157 -----H-----ENLLKVSOLENYKKYKEGL-QDLEGLKDVDTNFOE 1193
QY 843 YSKTQ-----HSAKAEHMQK-KRRKQENKD-----EALPPPPPPFPFAPFASVINL 886
Db 1194 LSKKHRLDTFNHESLLRQSASVKEKLSLASSENKDLNKNVSLTKOVNELSPKASKVPEL 1253
QY 887 LNCIDIMYILRVFERAIDTDSNLMWTEGMLQMAFHIALGLLEKQOQKAPKEEYTFDF 946
Db 1254 -----ERKI-----TNLMHEYSQLGKTFEDEKRRKAL 1279
QY 947 YHKASRLGSSAMNIQMLEKLGIPQLECGKDMITWILQMFDTVVKRLREKSLIVATTSG 1006
Db 1280 I---ASRDNEELSLKSELESKRKL-EVEYQK-----VLEEVKTRSLRSEVTLIRNKVAD 1331
QY 1007 SESIKNDEITHDKAEARKKAEARLHRQKIMAOMSALQKNFIETHKLMYDNTSEMPGK 1066
Db 1332 HESIRS-KLSEVEMKLVDRK-----ELNSALDSCKKREAEIHRUKEHRPS---GK 1378
QY 1067 EDSIMEEESTPAV 1079
Db 1379 ENNIPAVKTEPV 1391
RESULT 13
ID CENE_HUMAN
AC Q02224:
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Centromeric protein E (CENP-E protein).
GN CENPE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
```

```
SEQUENCE FROM N.A.
MEDLINE=93024922; PubMed=14069711;
Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;
"CENP-E is a putative kinetochore motor that accumulates just before
mitosis.";
Nature 359:536-539(1992).
[2]
CHARACTERIZATION.
MEDLINE=95196755; PubMed=7889940;
Thrower D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;
"Mitotic HeLa cells contain a CENP-E-associated minus end-directed
microtubule motor.";
EMBO J. 14:918-926(1995).
[3]
CHARACTERIZATION.
MEDLINE=98437347; PubMed=9763420;
Chan G.K.T., Schaar B.T., Yen T.J.;
"Characterization of the kinetochore binding domain of CENP-E reveals
interactions with the kinetochore proteins CENP-F and hBUBR1.";
J. Cell Biol. 143:49-63(1998).
RL J. Cell Biol. 143:49-63(1998).
CC -1- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
KINETOCHORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE
OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
AND/OR SPINDLE ELONGATION.
CC -1- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.
CC -1- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCHORES DURING
CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS
QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.
-----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; Z15005; CAA78727.1;
PIR; S28261; S28261.
HSP; P17119; 3KAR.
Genew; HGNC:1856; CENPE.
GK; Q02224;
MIM; 117143;
GO; GO:0005699; C:kinetochore; TAS.
GO; GO:0005634; C:nucleus; TAS.
GO; GO:0008350; F:kinetochore motor activity; TAS.
GO; GO:0000067; P:DNA replication and chromosome cycle; TAS.
GO; GO:0000707; P:mitotic chromosome movement; TAS.
GO; GO:0007080; P:mitotic metaphase plate congression; TAS.
InterPro; IPR001752; Kinesin_motor.
Pfam; PF00225; Kinesin; 1.
PRINTS; PR00380; KINESINHEAVY.
SMART; SM00129; KISC; 1.
PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;
Cell cycle; Centromere.
DOMAIN 1 335 KINESIN-MOTOR.
DOMAIN 336 2471 COILED COIL (POTENTIAL).
DOMAIN 2472 2663 GLOBULAR (POTENTIAL).
FT NP_BIND 86 93 ATP (BY SIMILARITY).
SQ SEQUENCE 2663 AA; 312087 MW; CEFCL3880C8CB8 CRC64;
Query Match 1.9%; Score 175.5; DB 1; Length 2663;
Best Local Similarity 19.4%; Pred. No. 0.028;
Matches 228; Conservative 189; Mismatches 406; Indels 355; Gaps 55;
QY 179 IKNSRCPLNEEVIVQARKIPSPVIVYVETIWEERKELPPEL-----QIREKNERYC 233
Db 733 VEENE--ALREVI-----LLSELKSLPSEVERLKRQKSEEL-- 770
```

Qy	234	VLFNDEHHSYDHYIYSQRLDCELABAQLHTTAIDKEGRRAVKAGAYACQAEKEDI	---	291
Db	771	-----HIITS-----EKDKLSEFVWHKESR	---VQGLLEEIGTKDDLAT	807
Qy	292	-KSHSENVSOHLHVEVLHSEIMAHOKFALRLGSWMN-KIMSYSDDPROQLFQCOACLREEP	349	
Db	808	TQSNYKSTQDFQNFKTLHMDFFQKYKWVLEENRNMQEIUNLSKEAQKF	-----LIFSFF	857
Qy	350	DSENPCLISRLMLWDARLYKGARKI	---LHE-----	383
Db	858	DSSLGALKTSELYKTOELQEKTRVQERLNMEOQLNEOLNDRSDPLQTVREKTLITEKL	917	
Qy	384	FMENYEKKLFAMEVKVYKOLQK	-----EYISDDHDR-SISITALSVMQFTVPTLARH	435
Db	918	QOTLEEVKTLTQEB-KDBKLQLOESQIERTDQLKSDIHTVMNIDTQEOURNALLESILKQH	976	
Qy	436	LIEBQNVISVITETLLEVLPE	-YLDNRNKNFQGYSDQKLGRVYAVICDLKYLLISK	491
Db	977	-----QETINTLKSKISEVSRNLHMEBNTGETKDFQCKMVG	-----IDKQDLLEAKNT	1026
Qy	492	PTIWERLNRQFEGFRSFKILITCMOGMEIEIRQVQGHLEVPDWEAAIAIOMQLK	---	548
Db	1027	QTLTADVKDNEIIBQOR	---KIFSLIOEKNELOQML-----	1073
Qy	549	--NILLMFQEWCADEELLIVA	-YKBECHAVNRCSTSFSSSKTVVQSGC-----	598
Db	1074	KENTEMTIENT	---QEBRLGLDELKQOEIVAQEKHAIKKEGELSRTCRLAEAVEEKL	1129
Qy	599	ETKSYRYSE	---DLVSTHPLS-----	637
Db	1130	KEKSOQLOEQKQQLLNQOEMSEMOKKINEIENLKNELKNKELTLEHMETERLELAQKLN	1189	
Qy	638	EFVSFEDQFQVELVYELPRLCLVLVAOVVAEMWRNRGLLSQVFFYQDVVKCREEMVDKDI	697	
Db	1190	E-NYEE	-----VKSITKE---AKVLKELQKSFETERDHLRGVIRIEA	1228
Qy	698	IMLQIGASLMDPNKFLLLVLRQY	-ELAEAFNKTISTK-----	750
Db	1229	TGLTKHEEL	---KIAHILKHEQETIDELRRSVSEKTAQIINTQDLKESHTKLOEB	1281
Qy	751	VLIYIVGERVYPGVNVTKEVIMREIHL	-----LCIEPMPHSAIAKNLPENE	799
Db	1282	IPVLHSEQELLPNVYKVKSETQETMNELELLELLETSQSTKDTTLARIEMERLRLNEKFKQESQ	1341	
Qy	800	NNETGLENVINKVATFKPGVSGHGVYELKDESILDFNMVFIYH	---YSKTOHSAKAEHMQ	855
Db	1342	EEIKSLTKERDNLTIKK	-----ALEVHKDQKE-----	1388
Qy	856	KRRKQENKOBALPPPPPPFCFAPFSKVINLLNCDIMMYIL	-----RTV-----	899
Db	1389	SLNMKXKDNETTKIVSEMEQFKPKDSA	---LLRIEIMGLSKRLQESHDEMKSVAKEKD	1445
Qy	900	---FERAIDTDSNLTWTEGMLQ	-MAFHILALGLLEEQQLOKA-----	949
Db	1446	DLQRLQEVLOQESDQLKENIKIYAKH	-----LETEEELKVACHCLKEQOEEETINELRVN	1499
Qy	950	ASRLGSSAMNIQMLLE	-----KLKGIQLEQCKDMITWILQMFTVKRLREKSKLIV	1001
Db	1500	LSEKETELSTIQOLEAINDKLQNKIQOIEIYEKESQLN	-IKOISEVQNVNMLQKFKDHRK	1558
Qy	1002	ATTSGSESIKND	-EITHD-----	1040
Db	1559	AKDSALQIESKMLELNNRLOESOEETIOMIKEEMKRVQEOALQIERTDQIKENTKEIVA	1618	
Qy	1041	QMSALQ	-----KNPIETHKL-----	1060
Db	1619	KMKESQEKQYQFLKMTAVNETQKMCETIEHLKEQFETQKLNLENIETENIRLTQILHENL	1678	
Qy	1061	SEMPG	---KEDSTMEESTPANSVDYRIALGPGRGFSVTEKEVLTICILQEOBEQEVKI	1114
Db	1679	EMRSVTKERDDLRDSVETELKVRDQKLENL	-----RETITBDL-EKQEBELKIVH	1727
Qy	1115	---ENNAMV	-LSACVOKSTALTQHRGKPIELSGEAL	1146

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AJ237839; CAB41415.1; -
 EMBL; AP000692; BAA89431.1; -
 EMBL; AP000692; BAA89431.1; JOINED.
 EMBL; AP000690; BAA89431.1; JOINED.
 EMBL; AP000691; BAA89431.1; JOINED.
 EMBL; AP001725; BAA95548.1; -
 EMBL; AB023150; BAA76777.2; -
 Genew; HGNC:1291; C21orf5.
 MIM; 604803; -
 GO; GO:0007275; P:development; NAS.
 Pfam; PF04118; Dopey_N; 1.
 CONFLICT 498 498 Q -> H (IN REF. 1).
 CONFLICT 1118 1118 G -> C (IN REF. 2 AND 3).
 CONFLICT 1149 1149 H -> P (IN REF. 2 AND 3).
 CONFLICT 1318 1318 L -> P (IN REF. 1).
 CONFLICT 1886 1886 V -> A (IN REF. 1).
 CONFLICT 1974 1974 E -> G (IN REF. 1).
 CONFLICT 2139 2139 G -> E (IN REF. 4).
 CONFLICT 2154 2154 K -> R (IN REF. 1).
 SEQUENCE 2298 AA; 258221 MW; 2C449AFE550739FO CRC64;

Query Match 1.9%; Score 170; DB 1; Length 2298;

Best Local Similarity 18.3%; Pred. No. 0.05;
 Matches 288; Conservative 201; Mismatches 551; Indels 536; Gaps 70;

QY 122 IDPTCVLCMDFQDSVHKHRYKMTSTGGFCDCGDTAEW----KTG--PFCVNEPGR 175
 DB LEFVILLQ-----PKQTSIHCLKQENAD--DLHRFNKKTSPREACAVPEQE 1047
 QY 176 AGTIKNSRCPLEEVIVQARKFPSPVYKVVEMTIWEEKEKLPELQIREKNERYCVL 235
 DB 1048 SGS---EEHPLSQFTVDREALVAEVEK-----EPEKPYLRGELSEELPYVEL 1095
 QY 236 FNDEHSHDVHYISLQALDCELAELAHHTAIDKEGRRAVAGAAOEAKEDEKSHS 295
 DB 1096 PRTAAG-----APDSSEHTSADTSSGH-----TDS 1122
 QY 296 ENVSQHLHVEVLHSEIMAHQFALRLGSMNKNINSYSDFRQIFCQACLREPDSENPC 355
 DB 1123 ENTS-----SFSPSHD-----LQELSNENCC 1145
 QY 356 LISRLMLDAKLYKGARKILHELIFSSFFEMEYKKLFAMEFVKY--YKOLQKEYISDDH 413
 DB 1146 -----APTHMGGRAPKRSALLAFAQSFSAKAKLSLVYVDSKTOASESFSSDE 1196
 QY 414 DRISITALSVMQFTVPTLARLHIEPQNVISVITETLLEVLPEYLDNRNKNFQGYSDQK 473
 DB 1197 EADLELQAL-----TTSRLKQOREQAEVAFKHLILYLPYDSR----- 1238
 QY 474 LGRVAVICDLKYLISKFTIWTERRMQLGFRSFLKILTCMQMEEIRROVGQHIEV 533
 DB 1239 --RVLYAFSVLEAVLKNP-----KEFIEAVSRSTSDTSTAHLNLSINLARH--- 1285
 QY 534 DPDWEAATA-----IQMLKNILLMFQEWACDEELLV-----AYKEC--- 572
 DB 1286 ---QCALIGQSYGKLQTOVPNV-----CPHSLLELLTYLCLSLFRSYPCYLKV 1333
 QY 573 -HKAYMRCSTSPISSSKTVVQ-----SCGHSLETKSYRVSDELVS-----IHLPLSR 618
 DB 1334 SHRDILNRDQVQKSVLELIRIMMQLVSVAKSSECKNVEFIHSLQRCVKQOEVLVLSL- 1392
 QY 619 TLAGLHVLRLSRIGAVSRHL-----EFVSEDF-----QVEVLVEYPLRC-LVLVAQV--- 664
 DB 1393 --ASWYTSQKRYGLATAHGRALPEDSLPEESLINLGQDIWSEHPLQIETLLKLLQVLIV 1450
 QY 665 -----VAEMWR--NGLSLISQVFFYQD-----VKCREEM 692
 DB 1451 LEHLGRAHEEAENOPDLRSREWRALNFQQAISALQYVQPHPLTSQGLLVSAYVRGLQPA 1510

QY 693 YDKDI-----IMLOIGASLMDPNKFKELLVLLQVYELAEAFN 727
 DB 1511 YGYCMHPAWSLVTHSLDPYFKSGLWTVTFVQVQCKNLDD-----LVKOYE-SESVK 1562
 QY 728 KTIST--KDQDLIKOYN--TLIEEMLQVLIYIVGERYVPGVGNVTKEEVTMRRIIHLCLIE 784
 DB 1563 LSVSTTSKRENISPDYPLTLEGLTTISHFCLLEQ-----ANQNKTTAAAGDPANL--- 1613
 QY 785 PMPHSATAKNLPENENNETGLENVINKVATFKKPGVSGHGVYVELKDESLKDNMMFYHVS 844
 DB 1614 RNARNAILEELPRTVNTMALLWNVLRKEETOKRP-----VDLLGATKGSSSVYF-KTT 1665
 QY 845 KTOHSHK-----AEHMOKKRKKOENKORALPPPPPEF---CP 878
 DB 1666 KTIQKILDFLNPHTAHLGVQLTAAVAANWRSKKAQRHSMKLIITPASASQLTLDVLC- 1724
 QY 879 AFSKVINLLCNDIMMYILRTVFER-----AIDTDSNLWTEGLMAFHIL----- 923
 DB 1725 ----ALSTLQTDTLHLVKEVVKRPPQVKGDEKSPLDIPVLQFCYAFQRLPVPALQE 1780
 QY 924 ----ALGELLEKQOLQKAPEEVTFDYHKASRLGSSAMNIOMLEKLGIPQLEQKDM 979
 DB 1781 NFSLLGVKESVQLNLAPP-----GYFLLSLMNDVTRTPNLENKKDQ 1825
 QY 980 ITWILQMFDTVKRLREKSLIVATSGSESIKNDIETHDKKAEKRKAARLHRQKTM 1039
 DB 1826 -----KDLQETQKILEAVNIAGSS-----L 1847
 QY 1040 AQMSALOKNFTETHKLMYDNTSEMPGKED-----SIWEESTPAVSDYSRIALGPKRG 1092
 DB 1848 EQTSWLSRN-LEVKAQQAQASLESDAEEDLDYDAAAASAMVSSASVSVQALSL----- 1901
 QY 1093 PSVTEKEVLTCIL-----COEQEVKTEN-----NAMVLSACVQKSTA 1130
 DB 1902 ----LAEVLASLLDMVYRSDEKAVPLISRLLYVFFVLYRNHSAYNAPSFRAGALLSS 1957
 QY 1131 LTQHRGPIELSGALDPLFMDPLAYCTYTGSCGHVMHVCQKYFEAVQLSSQRIHV 1190
 DB 1958 LSGYAYTKRAMRKVLE-LFLDP--AFFQMDTSCVH-----WKSIIIDL-LTHEKTMFK 2007
 QY 1191 DLPLESG-----EVLCPCKSLCNTVIP-----IIPLOKIN-----SENAD 1229
 DB 2008 DLMNQSSSKLFSFQKAMLLRQAPAFSGELDQYHLYLPLQIERTDNLVRQOTS 2067
 QY 1230 ALAQLLTLARQIQLARISGYNIRHAKGNPIPT-----FNGQMGGSTLEFHSILSFG 1284
 DB 2068 VAAQMFLEFR--VLLLRISP--QHLTSLMFWIMVSELIQFTQLEEDLKDEDESURSTN 2121
 QY 1285 VESSIKYS-----NSIKE-----MVLFTATTYIRIGLVKVPDPDERPRVPMLTWSTCAF 1332
 DB 2122 KVNRTKVSVPDANGPSVGEIPOSELILYLSACKFLDTALSPFPD---KMLPFIQYVWAF 2177
 QY 1333 TIQAIENLGDGKPLFGALONRONGKALKMQFAVORITCPQVLIQKHVRLLSV--- 1390
 DB 2178 IPEY-----DTEGPAFLSDVEENHQEK-----PHTVRIELEKL 2212
 QY 1391 -LPNIKSEDPCLLSIDLFLVLGAVLAFPSLDVDPDVLQDPSSVSSVNHLYLPHLITM 1449
 DB 2213 KFGIISSEDEITMKS-----EFPL-----LRQHSVSSIRQLMPPFTWTLNG 2252
 QY 1450 AHMLQILLTVDLT-GLP 1464
 DB 2253 AFKTQRLPADSPGTP 2268

RESULT 15

AKA9 HUMAN

ID AKA9_HUMAN STANDARD; PRT: 3911 AA.
 AC Q99956; O14869; O43355; O94895; Q9UQH3; Q9UQ04; Q9Y6B8; Q9Y6Y2;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)


```
FT DOMAIN 1434 1459 COILED COIL (POTENTIAL).
FT DOMAIN 1585 1659 COILED COIL (POTENTIAL).
FT DOMAIN 1857 2455 COILED COIL (POTENTIAL).
FT DOMAIN 2544 2561 COILED COIL (POTENTIAL).
FT DOMAIN 2603 2776 COILED COIL (POTENTIAL).
FT DOMAIN 3065 3092 COILED COIL (POTENTIAL).
FT DOMAIN 3124 3470 COILED COIL (POTENTIAL).
FT DOMAIN 3587 3689 COILED COIL (POTENTIAL).
FT DOMAIN 3726 3730 POLY-LEU.
FT DOMAIN 203 292 GLN-RICH.
FT DOMAIN 321 1010 GLU-RICH.
FT DOMAIN 1846 2772 Missing (in isoform 2 and isoform 3).
FT VARSPLIC 17 28 /FTid=VSP_004102.
FT VARSPLIC 1637 1642 QLOEEI -> LATRD (in isoform 4).
FT VARSPLIC 1643 3911 Missing (in isoform 4).
FT VARSPLIC 2175 2182 /FTid=VSP_004104.
FT VARSPLIC 2175 2183 Missing (in isoform 3).
FT VARSPLIC 2175 2183 /FTid=VSP_004105.
FT VARSPLIC 2895 2907 /FTid=VSP_004106.
FT VARSPLIC 2895 2907 VGFYNNCFSTLC -> GSSIPELAHSDAYOTREICSS
FT VARSPLIC 2895 2948 Missing (in isoform 5).
FT VARSPLIC 3901 3911 /FTid=VSP_004108.
FT VARSPLIC 3901 3911 STQPHAGMRR -> ALSUTTSWOHHSARPTAPLFFELSH
FT VARSPLIC 3901 3911 SLG (in isoform 6).
FT VARSPLIC 3901 3911 /FTid=VSP_004109.
FT VARSPLIC 3901 3911 K -> QO.
FT VARSPLIC 3901 3911 /FTid=VAR_010926.
FT VARSPLIC 3901 3911 E -> Q (IN REF. 3).
FT VARSPLIC 3901 3911 M -> I (IN REF. 3).
FT VARSPLIC 3901 3911 E -> G (IN REF. 3).
FT VARSPLIC 3901 3911 R -> S (IN REF. 3).
FT VARSPLIC 3901 3911 R -> S (IN REF. 3).
FT VARSPLIC 3901 3911 H -> N (IN REF. 3).
FT VARSPLIC 3901 3911 K -> N (IN REF. 3).
FT VARSPLIC 3901 3911 QKH -> PRP (IN REF. 1 AND 2).
FT VARSPLIC 3901 3911 Q -> P (IN REF. 1 AND 2).
FT VARSPLIC 3901 3911 Q -> P (IN REF. 1 AND 2).
FT VARSPLIC 3901 3911 V -> D (IN REF. 3).
FT VARSPLIC 3901 3911 V -> E (IN REF. 3).
FT VARSPLIC 3901 3911 R -> P (IN REF. 1 AND 2).
FT VARSPLIC 3901 3911 N -> T (IN REF. 3).
FT VARSPLIC 3901 3911 V -> G (IN REF. 3).
FT VARSPLIC 3901 3911 MISSING (IN REF. 5).
FT VARSPLIC 3901 3911 A -> P (IN REF. 3).

Query Match 1.8%; Score 168.5; DB 1; Length 3911;
Best Local Similarity 18.7%; Pred. No. 0.14;
Matches 201; Conservative 165; Mismatches 364; Indels 347; Gaps 50;

QY 208 EMTIWEKEKLPPELQIRKERNRYCVLFNDEHSHDYHVIYSQRLDCELABAQLHTTA 267
Db 728 EMTL--QINELQKEIETILRQEEK-----EKGTEQEVEQLQ--LKTELLEKOMK--- 772
QY 268 IDEGRRVAKGAYACQAEKEDIKSHSENVQHPHVLHVEHSEIMAHQKALRLGSMWN 327
Db 773 -ERENDLQKFAQLAENSILDKERTLEDMLK-----IHTPVSOEERLI-----FLD 819
QY 328 KIMSYSSDP---RQIFCQACLREPDSENPCLISRLMLMDAKLYKGARKILHELIFSSFF 384
Db 820 SIKSKSDSWKEI--ELIENEDLKQCCI-----QINEEIEKQRTNFFSAEKN 868
QY 385 MEMEYKKLFAMFVYKOLQKEY-----ISDHDHSISITALSVQMTVPTLARHLIEE 439
Db 869 FEVN-----YQELQEEYACLLKVKDDLEDKSN-----KQLELEY 901
QY 440 QNIVSIVITETLLEVPYLDNRN-----KFNQGYSDKLGRRYAVICDLKYLISKPTIW 495
Db 902 KSKLKALNEEL-----HLQRIINPTVKMKSSVFDEDK--TFVAETLEMEGVVEKDTTEL 953
```

```
QY 496 TERLRMQFLGFRSFLKILTCMQGMBEIRRRVQO--HIEVDPDWEAAIAIATOMQLKNILLMF 554
Db 954 MEKLEVTKREK-----LELSORLSDEQLKQKHGEISFLNEEVKSLKQKEQVSLRC 1006
QY 555 QEW-----CACDEE-----LLLVAYKECHKAVMRCSTSFSSSTVTVQSCG 595
Db 1007 RELEIIINHNRANVQSCDTQVSSLLDGVVTTSRGAESGVKNVKSFEESKIMWED-- 1064
QY 596 HSLKTSYRVSED-----LVSIHLPLSRTLGLH-VRLSRLGAVSRLHFEVSPEDFQVEV 649
Db 1065 -KVSFENNTVGEESKOEQIILDHLSVTKESSLRATQPSENDKLQKLNVLKSEQNDLRL 1123
QY 650 LVEPLRCLVLVAQVVAEMRRRNGLSLSQVYYQDVVKREMYDKDIIMLQIGASIMDP 709
Db 1124 QMEAQRICLSLVYS-----THVDQVREYME-----NEKDKALCSLKEELIFA 1165
QY 710 NKELLVLQR-YELAAEFNKTTISTKDDQ-----LIKQYNTLIEE-----MLOVLIYTVGER 759
Db 1166 QEEKIKELQKIHQLELOTMKTQETGDEGKPLHLIGLKQKAVSEBSCSYFLQTLCSVLGEY 1225
QY 760 YVPGVGNVTKBEVTMRIIHLICIEBPMPSALAKNLPENNETGLENVINKVATEKKPG 819
Db 1226 YTPAL-----KCEV-----NAEDKEN----- 1241
QY 820 VSGHGVVELKDESLEKDFNMFYHYHYSKTHSKAEHMOKRRKROENKDEALPPPPPPFCPA 879
Db 1242 -SGDYISENDEPELDQRYEVQDFQENMHT-----LLNKVTEEYNK----- 1281
QY 880 FSKVINLLNCDIMVILRTVFERADTDSNLW---TEGMLQMAFHIALGLLLEEKQOLQK 936
Db 1282 -----LLVLQTRL-----SKIWQOOTDGM-----KLEFGE-----N 1308
QY 937 APEEEVTF-----DFYHKASRLGSSAMNIOMLLLEKLGIPQLEGOKMITWI 983
Db 1309 LPKEETEFLSIHSQMTNLEDIDVNHK-----SKLSSLODLEKTKLEEQVQVELESLSL 1362
QY 984 LQMFDTVKRLREKS--CL-----IVATTSGSESIKNDIETHDKKAEARKKAEAAHLRQ 1036
Db 1363 QOOLKETEQNYEAETHCLQKRLQAVSESVPPSLPVDVSVITESDA-----Q 1409
QY 1037 KIMAQMSALQKNFTETHKLMYDNTSEMPGKEDSIMEEESTPAVSDYSRIALGPKRGPSVT 1096
Db 1410 RTMYPGSCVKKNI-----DGTIEFSG-EFGVKEETNI-----VKLL 1444
QY 1097 KEVLTCTLCOEQEVKTENNAMVLSACVQKSTALTQHRGKPIELSGEALDPLFMDPDLA 1156
Db 1445 EKQY-----QEQLLEEVAKVIVSMSIAFAQOTELSR-----ISG----- 1478
QY 1157 YGTYTSGCGHYMHAVCWOK--YFEAVQLSSQORIHVDLF---DLESGEYLCPLCKSL 1208
Db 1479 -GKENTASSQAHAVCQEQEHYFNEMKL-SQDQIGFQTFETVDVVKFKEEFKPLSKEL 1533
```

Search completed: September 25, 2003, 14:55:22

Job time : 50 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 25, 2003, 14:51:51 ; Search time 148 Seconds
(without alignments)
3023.400 Million cell updates/sec

Title: US-09-724-126a-19
Perfect score: 9141
Sequence: 1 AMEGNMADEAGGTERMEIS.....ETARSQETNQMLFGFNQALL 1734

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_23:*
1: sp-archaea:*
2: sp-bacteria:*
3: sp-fungi:*
4: sp-human:*
5: sp-invertebrate:*
6: sp-mammal:*
7: sp-mhc:*
8: sp-organelle:*
9: sp-phage:*
10: sp-plant:*
11: sp-rodent:*
12: sp-virus:*
13: sp-vertebrate:*
14: sp-unclassified:*
15: sp-rvirus:*
16: sp-bacteriap:*
17: sp-archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9087	99.4	1749	4 Q8IYW7	Q8IYW7 homo sapien
2	8791.5	96.2	1709	4 Q8IYW6	Q8IYW6 homo sapien
3	8480	92.8	1757	11 Q70481	Q70481 mus musculus
4	4262	46.6	849	11 Q8BN40	Q8BN40 mus musculus
5	4229	46.3	1755	4 Q8IYW8	Q8IYW8 homo sapien
6	4209	46.0	811	4 Q96JY4	Q96JY4 homo sapien
7	4184	45.8	1755	11 Q8CGW0	Q8CGW0 mus musculus
8	3062	33.5	1275	4 Q15057	Q15057 homo sapien
9	3053	33.4	1273	4 Q9NU68	Q9NU68 homo sapien
10	2593.5	28.4	1109	11 Q8K2I6	Q8K2I6 mus musculus
11	2232.5	24.4	861	11 Q8BUL9	Q8BUL9 mus musculus
12	2210.5	24.2	1824	5 Q8SX71	Q8SX71 drosophila
13	2208.5	24.2	1824	5 Q9VX91	Q9VX91 drosophila
14	1886.5	20.6	1927	5 P91133	P91133 caenorhabdi
15	1716	18.8	333	4 Q60708	Q60708 homo sapien
16	1692	18.5	329	4 Q75492	Q75492 homo sapien

17	1688	18.5	777	11 Q8ROV7	Q8ROV7 mus musculus
18	1067	11.7	229	11 Q8C5K3	Q8C5K3 mus musculus
19	911.5	10.0	410	11 Q8RI30	Q8RI30 mus musculus
20	606.5	6.6	1225	10 Q9LZ95	Q9LZ95 arabidopsis
21	598	6.5	212	4 Q9H578	Q9H578 homo sapien
22	535.5	5.9	1872	3 Q07963	Q07963 saccharomyc
23	513	5.6	2153	5 Q19330	Q19330 caenorhabdi
24	470.5	5.1	1264	10 Q94JL2	Q94JL2 oryza sativ
25	362	4.0	1470	5 Q9W3M3	Q9W3M3 drosophila
26	359	3.9	2230	5 Q8NMN0	Q8NMN0 dictyosteli
27	304	3.3	731	5 Q9W3M5	Q9W3M5 drosophila
28	303	3.3	1432	5 Q8IRN8	Q8IRN8 drosophila
29	292	3.2	795	10 Q9LE94	Q9LE94 arabidopsis
30	288.5	3.2	5439	5 Q8I586	Q8I586 plasmodium
31	284.5	3.1	794	10 Q9G248	Q9G248 arabidopsis
32	257	2.8	437	4 Q8IWE7	Q8IWE7 homo sapien
33	252	2.8	343	4 Q8ND96	Q8ND96 homo sapien
34	208	2.3	378	11 Q8CAW0	Q8CAW0 mus musculus
35	205.5	2.2	3187	11 Q63714	Q63714 rattus norv
36	196	2.1	280	10 Q94JL1	Q94JL1 oryza sativ
37	189.5	2.1	8749	4 Q8NF91	Q8NF91 homo sapien
38	186	2.0	1410	4 Q14221	Q14221 homo sapien
39	185	2.0	1449	6 Q9BG87	Q9BG87 bos taurus
40	184	2.0	1411	4 Q15075	Q15075 homo sapien
41	184	2.0	1979	4 Q95949	Q95949 homo sapien
42	182.5	2.0	2442	4 Q9H450	Q9H450 homo sapien
43	182	2.0	2442	4 Q14812	Q14812 homo sapien
44	181.5	2.0	2442	4 Q60588	Q60588 homo sapien
45	181.5	2.0	2579	11 Q9ESD3	Q9ESD3 mus musculus

ALIGNMENTS

RESULT 1

Q8IYW7 PRELIMINARY; PRT; 1749 AA.
ID Q8IYW7
AC Q8IYW7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Ubiquitin ligase E3 alpha-I.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Han H.Q., Kwak K.;
RT "Full-length human ubiquitin ligase E3 alpha-I (E3 alpha).";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY061886; AAL32103.1;
KW Ligase.
SQ SEQUENCE 1749 AA; 200210 MW; 3AE0E1A749884971 CRC64;

Query Match 99.4%; Score 9087; DB 4; Length 1749;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1728; Conservative 0; Mismatches 1; Indels 20; Gaps 1;
Qy 6 MADEAGGTERMEISAEELPQTPQRLASWMDQVDFYTAFLHHLAQLVPEIYFAEMDPDL 65
Db 1 MADEAGGTERMEISAEELPQTPQRLASWMDQVDFYTAFLHHLAQLVPEIYFAEMDPDL 60
Qy 66 QKESVQMSFTPLEWYLFGEPPDPCLEKHKHSAFQLCGRVFKSGGTTYSRCDCAIDPT 125
Db 61 QKESVQMSFTPLEWYLFGEPPDPCLEKHKHSAFQLCGRVFKSGGTTYSRCDCAIDPT 120
Qy 126 CVLCMDCQFQSVHKNHRYKMHSTGTGGFCDCGDTFAWKTPFCVNHPEGRAGTIKENSRC 185
Db 121 CVLCMDCQFQSVHKNHRYKMHSTGTGGFCDCGDTFAWKTPFCVNHPEGRAGTIKENSRC 180
Qy 186 PLNEEVTVQARKIPSPSVIKYVWETIWEKEELPELQIREKNERYVCVLFNDEHSHSDH 245

Db 181 PLNEEVIVQARKIPSPVIVKVVWMTIWEEBKELPPELOIREKNERYCYVLFNDHSHSYDH 240
Qy 246 VIYSLQALDCELAELAQHTTAIDKEGRRAVYAGAYAACQAKEDIKSHSNVSOHPHUH 305
Db 241 VIYSLQALDCELAELAQHTTAIDKEGRRAVYAGAYAACQAKEDIKSHSNVSOHPHUH 300
Qy 306 EVLHSEITMAHOKFALRLGSMWNTMSYSSDFRQIFCQACLEEDPDPSENPCILSLRLMLWA 365
Db 301 EVLHSEITMAHOKFALRLGSMWNTMSYSSDFRQIFCQACLEEDPDPSENPCILSLRLMLWA 360
Qy 366 KLYGARKILHELLFSFFMEMEYKILFAMEFVYKQLOKEYISDDHDSISITALSQV 425
Db 361 KLYGARKILHELLFSFFMEMEYKILFAMEFVYKQLOKEYISDDHDSISITALSQV 420
Qy 426 MFTVPTLARHLIEBQNVISVITETLLEVLPEYLDNRNKNFQGYSDQKLGRVYAVICDLK 485
Db 421 MFTVPTLARHLIEBQNVISVITETLLEVLPEYLDNRNKNFQGYSDQKLGRVYAVICDLK 480
Qy 486 YILISKPTIWTIRLMOFLEGFRSLKILTCMQGMEETRRQVGHIEVDPDWEAAIAIQM 545
Db 481 YILISKPTIWTIRLMOFLEGFRSLKILTCMQGMEETRRQVGHIEVDPDWEAAIAIQM 540
Qy 546 QLKNIILMFQWCACDELLAVAYEKCHKAVMRCSTSFISSTKTVQSCGHSLETKSYRV 605
Db 541 QLKNIILMFQWCACDELLAVAYEKCHKAVMRCSTSFISSTKTVQSCGHSLETKSYRV 600
Qy 606 SEDLVSIHLPLSRTLGLHVLRLSRLGAVSRILHEFVSFEDFQVLEVPLRCLVLAQV 665
Db 601 SEDLVSIHLPLSRTLGLHVLRLSRLGAVSRILHEFVSFEDFQVLEVPLRCLVLAQV 660
Qy 666 AEMWRRNGLSLSQVFFYQDYVKCREMYDKDIIMLOIGASLMDPNKFLLLVLQRYELAE 725
Db 661 AEMWRRNGLSLSQVFFYQDYVKCREMYDKDIIMLOIGASLMDPNKFLLLVLQRYELAE 720
Qy 726 FNKTIISTKQDOLIKOYNTLIBEMLOVLIYIVGERYVPGVGNVTKEEVTMREIHLCLTEP 785
Db 721 FNKTIISTKQDOLIKOYNTLIBEMLOVLIYIVGERYVPGVGNVTKEEVTMREIHLCLTEP 780
Qy 786 MPHSAIAKLNLENENNETGLNVINKVATFKPKPGVSGHGYELKDESKDFNMVYHYSK 845
Db 781 MPHSAIAKLNLENENNETGLNVINKVATFKPKPGVSGHGYELKDESKDFNMVYHYSK 840
Qy 846 TOHSAEAMHOKRRKQENKDALPPPPPPPCPAFSKVINLLNCDIMMYILRTVPERAID 905
Db 841 TOHSAEAMHOKRRKQENKDALPPPPPPPCPAFSKVINLLNCDIMMYILRTVPERAID 900
Qy 906 TDSNLWTGMLQMAFHIALGLLEKQOLOKAPPEEVTDFVHKASRLGSSAMNTOMLLE 965
Db 901 TDSNLWTGMLQMAFHIALGLLEKQOLOKAPPEEVTDFVHKASRLGSSAMNTOMLLE 960
Qy 966 KLKGIPOLEGOKDMITWILQMFDTVKRLRKSCLIVATTSGSESINKNDEITHDKEAERK 1025
Db 961 KLKGIPOLEGOKDMITWILQMFDTVKRLRKSCLIVATTSGSESINKNDEITHDKEAERK 1020
Qy 1026 RKAEEARLHROKIMQMSALOKNFETHTKLMYDNTSEMPGKEDSTMEESTPAVSDYSRI 1085
Db 1021 RKAEEARLHROKIMQMSALOKNFETHTKLMYDNTSEMPGKEDSTMEESTPAVSDYSRI 1080
Qy 1086 ALGPKRGPSTVEKEVLTICLQEEQEVKIENAMVLSACVOKSTALTOHRGKPIELSGEA 1145
Db 1081 ALGPKRGPSTVEKEVLTICLQEEQEVKIENAMVLSACVOKSTALTOHRGKPIELSGEA 1140
Qy 1146 LDPLFMDPDLAYGTGSCGHMVHACVQKTFEAVQLSSQQRHIVDLDFLESSEYLCPLC 1205
Db 1141 LDPLFMDPDLAYGTGSCGHMVHACVQKTFEAVQLSSQQRHIVDLDFLESSEYLCPLC 1200
Qy 1206 KSLCNTVPIIPILOPKINSNADALQLLTLARMIQTVLARISGYNIRHAKGNPIPIF 1265
Db 1201 KSLCNTVPIIPILOPKINSNADALQLLTLARMIQTVLARISGYNIRHAKGNPIPIF 1260
Qy 1266 FNOGMDSTLBFHSLSGVSSIKYSIKEMVTLFATTIYRIGLKVPPDPRVPM 1325
Db 1261 FNOGMDSTLBFHSLSGVSSIKYSIKEMVTLFATTIYRIGLKVPPDPRVPM 1320

Qy 1326 TWSTCAFTIQTATENLLGDEGKPLFGALQNRQHNGLKALMQFAVAQORITCPQVLIQKHLVR 1385
Db 1321 TWSTCAFTIQTATENLLGDEGKPLFGALQNRQHNGLKALMQFAVAQORITCPQVLIQKHLVR 1380
Qy 1386 LLSVVLPNKISBDTPCLLSIDLFLHVLGAVLAPPSLYWDDPVLQPPSSVSSYNHLYLRFH 1445
Db 1381 LLSVVLPNKISBDTPCLLSIDLFLHVLGAVLAPPSLYWDDPVLQPPSSVSSYNHLYLRFH 1440
Qy 1446 LITMAHMLQIILLTVDTGLTGLAQVQDESEEAHSASSFFAEISQYTSIGCDIPGWLWVS 1505
Db 1441 LITMAHMLQIILLTVDTGLTGLAQVQDESEEAHSASSFFAEISQYTSIGCDIPGWLWVS 1500
Qy 1506 LKNGITPYLRCAALFFHYLLGVTPPEELHTNSAEGYSALCSYLSLPTNLLFLFQEWYWD 1565
Db 1501 LKNGITPYLRCAALFFHYLLGVTPPEELHTNSAEGYSALCSYLSLPTNLLFLFQEWYWD 1560
Qy 1566 VRPILORRCADPALLNCLKOKNTVVRYPRKRNLSLIELPDDYSCLLNQASHFRCPRSADDE 1625
Db 1561 VRPILORRCADPALLNCLKOKNTVVRYPRKRNLSLIELPDDYSCLLNQASHFRCPRSADDE 1620
Qy 1626 RKHPVLCFCGAILCSQICQOIEVNGEVEGACIFHALHC----- 1665
Db 1621 RKHPVLCFCGAILCSQICQOIEVNGEVEGACIFHALHCAGVCIFLKTRECRVVLVEG 1680
Qy 1666 KARGCAYPAPYDVEYGETDGLKRGPNLHLSRERYRKLHLVWQOHCIEBIARSQETNQM 1725
Db 1681 KARGCAYPAPYDVEYGETDGLKRGPNLHLSRERYRKLHLVWQOHCIEBIARSQETNQM 1740
Qy 1726 LFGFNWQLL 1734
Db 1741 LFGFNWQLL 1749

RESULT 2
Q8IWY6 PRELIMINARY; PRT; 1709 AA.
AC Q8IWY6;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DE UBRL E3a ligase (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=224340442; PubMed=12434312;
RA Dgany O., Avidan N., Delaunay J., Krasnov T., Shalmon L., Shalev H.,
RA Eidelitz-Markus T., Kapelushnik J., Cattani D., Pariente A.,
RA Tulliez M., Cretien A., Schischmanoff P.O., Tolascio A., Fibach E.,
RA Koren A., Rossler J., Le Meirer M., Yaniv I., Zaizov R., Ben-Asher E.,
RA Olender T., Lancet D., Beckmann J.S., Tamary H.;
RT "Congenital dyserythropoietic anemia type I is caused by mutations in
RT codanin-1";
RL Am. J. Hum. Genet. 71:1467-1474 (2002).
DR EMBL; AF525401; AA014997.1; --
KW Ligase.
FT NON_TER 1 1709
FT NON_TER 1709 1709
SQ SEQUENCE 1709 AA; 195279 MW; B2B17231A1020FID CRC64;

Query Match 96.2%; Score 8791.5; DB 4; Length 1709;
Best Local Similarity 98.0%; Pred. No. 0;
Matches 1676; Conservative 1; Mismatches 12; Indels 21; Gaps 2;

Qy 23 LPQTPQRLASWMDQVDFYTAFLHLHLAQLVPEIFYFAEMDPLEKQESVQMSIFTPLEWY 82
Db 1 LPQTPQRLASWMDQVDFYTAFLHLHLAQLVPEIFYFAEMDPLEKQESVQMSIFTPLEWY 60
Qy 83 LFGEDPDICLEKLKHSAGFQLCGRVPKSGETTYSRCDADPTCVLCMDCFDQSVHKNHR 142

Db 61 LFGEEDPICLEKLKHSAGFQLGRVFKSGGETTYSRCDCAIDPTCVLCMDCFQDSVHKNR 120
QY 143 YKMHSTGCGFCGDTAEWKTPFCVNHPEGRAGTIKENSRCPLNEEVIVQARKFPSPV 202
Db 121 YKMHSTGCGFCGDTAEWKTPFCVNHPEGRAGTIKENSRCPLNEEVIVQARKFPSPV 180
QY 203 IKYVEMTWEBEKELPPELQIREKNERYCYVLFNDEHHSYDHVYISLQALDCELAEAQ 262
Db 181 IKYVEMTWEBEKELPPELQIREKNERYCYVLFNDEHHSYDHVYISLQALDCELAEAQ 240
QY 263 LHTTADKGRRAVAGAYAAQOEAKEDIKSHSENVSOHPLHVEVHSEIMAHQKFAIRL 322
Db 241 LHTTADKGRRAVAGAYAAQOEAKEDIKSHSENVSOHPLHVEVHSEIMAHQKFAIRL 300
QY 323 GSWMNKINSYSDRFQIFQOACLRPEPDSFNCLISRLMDAKLYKGARKIILHELIFSS 382
Db 301 GSWMNKINSYSDRFQIFQOACLRPEPDSFNCLISRLMDAKLYKGARKIILHELIFSS 360
QY 383 FFMENYKLFAMFVKYKQLOKEYISDDHRSISITALSVMFTVPTLARHLIEONV 442
Db 361 FFMENYKLFAMFVKYKQLOKEYISDDHRSISITALSVMFTVPTLARHLIEONV 420
QY 443 ISVITETLLEVLPEYLDNRNKNFQGYSDKLGVRVAVICDLKYIILISKPTIWERLRMQ 502
Db 421 ISVITETLLEVLPEYLDNRNKNFQGYSDKLGVRVAVICDLKYIILISKPTIWERLRMQ 480
QY 503 FLEGFRSFLKIULTCMQMEIEIRQVQGHTEVDPDNEAAIAIOMQLKNILLMFOEWCACDE 562
Db 481 FLEGFRSFLKIULTCMQMEIEIRQVQGHTEVDPDNEAAIAIOMQLKNILLMFOEWCACDE 540
QY 563 ELLLVAYKECHKAVMRCSTSFSSKTVVQSGHSLSTKSYRVEDLVSIHLPLSRTLGA 622
Db 541 ELLLVAYKECHKAVMRCSTSFSSKTVVQSGHSLSTKSYRVEDLVSIHLPLSRTLGA 600
QY 623 LHVRLSRLGAVSRLEHFEVSFDFQVEVLVEYPLRCLVLVAQVVAEMRRNGSLISQVFF 682
Db 601 LHVRLSRLGAVSRLEHFEVSFDFQVEVLVEYPLRCLVLVAQVVAEMRRNGSLISQVFF 660
QY 683 YQDVKCREMYDKDIIMIQIGASLMDPNKFLLLVLRQYELAEAFNKTISTKQDOLIKQYN 742
Db 661 YQDVKCREMYDKDIIMIQIGASLMDPNKFLLLVLRQYELAEAFNKTISTKQDOLIKQYN 720
QY 743 TLIEEMQLVLIYIGERYVPGVNTKEEVTWREIHLICIEPMPHSAITAKNLPENENNE 802
Db 721 TLIEEMQLVLIYIGERYVPGVNTKEEVTWREIHLICIEPMPHSAITAKNLPENENNE 780
QY 803 TGLENVINKVATFKKPGVSGHGYELKDESLKDFNMVYVYHYSKTOHSAEHMOKKRKOE 862
Db 781 TGLENVINKVATFKKPGVSGHGYELKDESLKDFNMVYVYHYSKTOHSAEHMOKKRKOE 840
QY 863 NKDEALPPPPPEFCPAFSKVINLLNCDITMYYILRTVFERAIDTDSNLTWTEGMQLQAFHI 922
Db 841 NKDEALPPPPPEFCPAFSKVINLLNCDITMYYILRTVFERAIDTDSNLTWTEGMQLQAFHI 900
QY 923 LALGLLEKQLOKQAPAEVETDFYHKASRLGSSAMNTIOMLEKLGIPQLEGQKDMITW 982
Db 901 LALGLLEKQLOKQAPAEVETDFYHKASRLGSSAMNTIOMLEKLGIPQLEGQKDMITW 960
QY 983 ILOMFDTVKRLREKSLIVATTSGSESIKNDEITHDKKAERKRAEARLHROKTIQAOM 1042
Db 961 ILOMFDTVKRLREKSLIVATTSGSESIKNDEITHDKKAERKRAEARLHROKTIQAOM 1020
QY 1043 SALQKNFTETHKLYMDNTSEMPGKEDSIWEESTPAVSDYSRIALGPKRGPSVTEKEVLT 1102
Db 1021 SALQKNFTETHKLYMDNTSEMPGKEDSIWEESTPAVSDYSRIALGPKRGPSVTEKEVLT 1080
QY 1103 CILQOEQEVKIENNAMVLSACVQKSTALTQHRGKPIELSGEALDPLFMDPDLAYGTYTG 1162
Db 1081 CILQOEQEVKIENNAMVLSACVQKSTALTQHRGKPIELSGEALDPLFMDPDLAYGTYTG 1140
QY 1163 SCGHVMAHVCWKQFEAVQLSSQQRHIVDLFDLESGEYLCPLCKSLCNTVPIPIIPLQPOK 1222
Db 1141 SCGHVMAHVCWKQFEAVQLSSQQRHIVDLFDLESGEYLCPLCKSLCNTVPIPIIPLQPOK 1200

QY 1223 INSENADALAQLLTLARWIOQTVLARISGYNIRHAKGENPIPIFFNQMGDSTLEPHSILS 1282
Db 1201 INSENADALAQLLTLARWIOQTVLARISGYNIRHAKGENPIPIFFNQMGDSTLEPHSILS 1260
QY 1283 FGVESSIKYSIKSEMVLIFATTIYRIGLKYPPDERDPRVPMLTWSTCAFTIOAENLIG 1342
Db 1261 FGVESSIKYSIKSEMVLIFATTIYRIGLKYPPDERDPRVPMLTWSTCAFTIOAENLIG 1320
QY 1343 DEGKPLFGALQNRHNGLKALMQFAVAQRITCPOVLIOKHLVRLLSVVLPNIKSBDTPCL 1402
Db 1321 DEGKPLFGALQNRHNGLKALMQFAVAQRITCPOVLIOKHLVRLLSVVLPNIKSBDTPCL 1380
QY 1403 LSIDLHVILGAVLAFPSLYWDDPVDLQPPSSVSSSYNNHLYLFHLITMAHMLQIILLTVDTG 1462
Db 1381 LSIDLHVILGAVLAFPSLYWDDPVDLQPPSSVSSSYNNHLYLFHLITMAHMLQIILLTVDTG 1440
QY 1463 LPLAQOVEDSEAHSSASFFFAEISQYTSIGSDIPGWYLVWSLKGITTPYURCAALRPH 1522
Db 1441 LPLAQOVEDSEAHSSASFFFAEISQYTSIGSDIPGWYLVWSLKGITTPYURCAALRPH 1500
QY 1523 YLLGVTPEELHTNSARGEYSALCSYLSLPTNLFLFOEYWDTPVPLLRRCADPALLNC 1582
Db 1501 YLLGVTPEELHTNSARGEYSALCSYLSLPTNLFLFOEYWDTPVPLLRRCADPALLNC 1560
QY 1583 LKOKNTVVRPRKRNLSIELPDDYSCLLNQASHFRCSADDERKHPVLCFCGAILCSQ 1642
Db 1561 LKOKNTVVRPRKRNLSIELPDDYSCLLNQASHFRCSADDERKHPVLCFCGAILCSQ 1620
QY 1643 NICCOEIVNGEEVCACIFHALHC-----KARGCAYPAPYLDEYGE 1682
Db 1621 NICCOEIVNGEEVCACIFHALHCAGVCIFLKIRECRVVLVEGKARGCAYPAPYLDEYGE 1680
QY 1683 TDPGLKRGNPLHLSRERYRKLHLVWQOHC 1712
Db 1681 TDPGLKRGNPLHLSRERYRKLHLVWQOHC 1709

RESULT 3

070481 PRELIMINARY; PRT: 1757 AA.
AC 070481: 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Ubiquitin-protein ligase E3 COMPONENT N-recognin (Ubiquitin-protein
ligase E3-alpha).
GN UBRI.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Kwon Y.T., Reiss Y., Fried V.A., Hershko A., Yoon J.K., Gonda D.K.,
RA Sanghan P., Copeland N.G., Jenkins N.A., Varshavsky A.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-446 FROM N.A.
RC STRAIN=129/SVJ;
RA Kwon Y.T., Reiss Y., Fried V.A., Hershko A., Yoon J.K., Gonda D.K.,
RA Sanghan P., Copeland N.G., Jenkins N.A., Varshavsky A.;
RT "The mouse and human genes encoding the recognition component of the
N-end rule pathway."
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF061555; AAC40165.1; -
DR EMBL; AF067379; AAC23678.1; -
DR EMBL; AF067371; AAC23678.1; JOINED.
DR EMBL; AF067372; AAC23678.1; JOINED.
DR EMBL; AF067373; AAC23678.1; JOINED.
DR EMBL; AF067374; AAC23678.1; JOINED.
DR EMBL; AF067375; AAC23678.1; JOINED.
DR EMBL; AF067376; AAC23678.1; JOINED.


```
Db 481 REEMYDKDIIMLQIGASLMDPNKFLLLVLRQYELAEAFNKTISTKDQDLIKQYNTLIEEM 540
QY 749 LQVLIYVIGERYVPGVGNVTKEVTWREIITHLILCIEPMPSATAKNLPENENNETGLENV 808
Db 541 LQVLIYVIGERYVPGVGNVTKEVTWREIITHLILCIEPMPSATAKNLPENENNETGLENV 600
QY 809 INKVATFKKPGVSGHGYELKDESLKDFNMVYHYHYSKTQSHKAEHMOKKRRKQENKDEAL 868
Db 601 INKVATFKKPGVSGHGYELKDESLKDFNMVYHYHYSKTQSHKAEHMOKKRRKQENKDEAL 660
QY 869 PPPPPPEFCAPSKVINLNCNDIMYILRTVFERAIDTDSNLTWTEGMLQMAFHIALGLL 928
Db 661 PPPPPPEFCAPSKVINLNCNDIMYILRTVFERAIDTDSNLTWTEGMLQMAFHIALGLL 720
QY 929 EKKQQLQKAPEREVTDFVHKASRLGSSAMNIOMLEKLGIPOLGOKDMITWILQMF 988
Db 721 EKKQQLQKAPEREVTDFVHKASRLGSSAMNIOMLEKLGIPOLGOKDMITWILQMF 780
QY 989 TVKRLREKSLIVATTSGSESINKNDIETHDK 1019
Db 781 TVKRLREKSLIVATTSGSESINKNDIETHDK 811

RESULT 7
Q8CGW0 PRELIMINARY; PRT; 1755 AA.
AC Q8CGW0;
DT 01-MAR-2003 (TremBrel. 23, Created)
DT 01-MAR-2003 (TremBrel. 23, Last sequence update)
DT 01-MAR-2003 (TremBrel. 23, Last annotation update)
DE Ubiquitin ligase E3 alpha-II (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/HeN;
RA Han H.Q., Kwak K.;
RT "Novel ubiquitin ligase E3 alpha-II.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY061885; AAL32102.1; -.
KW Ligase.
FT NON_TER
SQ SEQUENCE 1755 1755 199165 MW; 21DD9EDE8E281B39C CRC64;

Query Match 45.8%; Score 4184; DB 11; Length 1755;
Best Local Similarity 45.9%; Pred. No. 9.5e-283;
Matches 824; Conservative 335; Mismatches 531; Indels 104; Gaps 29;

QY 6 MADE---EAGGTER--MEISAEIPQTPQRLASWDOQVDFTAFLLHQAQLVPEIYFAEM 60
Db 1 MASEMEPEVQAIADRSLLCSAE-----EIAGRWLQATDLNREYVQHLAHCVPKIYCRGP 54
QY 61 DPDLKQESVOMSTFTPLEWYLFGEEDPDICLEKLKHSQ-AFOLGGRVFKSGTTTSCRD 119
Db 55 NFPPOKEDTLAGHLLGPNWEVICAEDPALGFPKLEQANKPSHLCGRVFKVGEPTTSCRD 114
QY 120 CAIDPTCVLCMDFODSVHKNHRYKMHTSTGGGFCDCGDTGTEAKWGTGFCVNHPEGRAGTI 179
Db 115 CAVDPTCVLCMCEFLGSIHRDHRYRMTTSGGGFCDCGDTGTEAKWKGYPYCKHKLSSSEVV 174
QY 180 -KENSRCPLNEEVIQVQARIFPFSVIKYVYVEMTIEWEEKELPPELQIREKNERYCYVLFND 238
Db 175 EEDPLVHLSVEDVIARTYNIFALMPRYAVDILTWEKESELPEDEVAEKSDDTYCYMLFND 234
QY 239 EHSYDHVYLSORALDCELABAQLHTTAIDKEGRRAVAKAGAYAAQCEAKEDIKSHENV 298
Db 235 EVHTTEQVYITLQKAVNCQKEAIGFATTVDRDRRPVRYGDFQYCDQAKTVIVRNTSRQ 294
QY 299 SOHPLHVEVLHSEIMAHQKFAIRLGSWMKNIMSYSSDFRQIFCOACLREPDSENCLIS 358
Db 295 TK-PLKVQVMHSSVAHQFGLKALSWLGSVIGYSDGLRILQCQVGLQEGPCDGENSLVD 353
```

```
QY 359 RLMLWDAKLYKGARKILHELIFSSFFMEMEYKKLFAMEFVKYKQLOKEYISDDHDSIS 418
Db 354 RLMLNDSKLWKGARSVYHOLFMSLMDLYKAKLFAIRFKNYROLQDFMEDDHERAVS 413
QY 419 ITALSVQMTPTPTLARHLIEEONVISTETLEVLPEYLDNRNNFNFOGYS---ODKLG 475
Db 414 VTALSVQFTPTPTLARMLLTLEENLMTVIKAFMDHL-KHRDAQGRQFERYTALQAFKFR 472
QY 476 RYAVVICDLKYLISKPTIWTERTLQMFLEGRFSRLKILTCMQGMBEIRRQVQHIIEVDP 535
Db 473 RVQSLILDLYKVLISKPTESWDELQKFLQGFDALELLKCMQGMPTITRQVQGHTEMEP 532
QY 536 DWEAIAIQOMOLKNILLMFOEWCACDEBELLVAYKECHKAVMRCSTSFSSSKTVVQS-C 594
Db 533 EWEAAFTLQMLKTHVISWQDWCALDEKVLIEAYKKCLAVLTQCHGSGFTDGGQPTLSIC 592
QY 595 GHSLETYSYRVSEDLVSIHLPLSRTLAGLHVRLSRLGAVSRSEHFVSPDFQFVEVLEYEP 654
Db 593 GHSVETIRYCVSQEKVSIHLPIISRLLAGLHVLLSKSEVAYKFPPELLPLSELPPMLIEHP 652
QY 655 LRCLVLVAQVVAEMWRNGLSLISQVFFYQDVVKREMYDKDIIMLQIGASLMDPNKFL 714
Db 653 LRCLVLCAQVHAGWMRRNGFSLVNGIYYHNVKCRREMFDKDI VMLQTGVSMMDPNHF 712
QY 715 LVLRQYELAEAFN-----KTISTK--DQDLIKQYNTLIEEMLOVLIYVIGERYVPGVGNV 767
Db 713 IMLSRFELYQLFSTPDYKGRFSSEVTHKDVVQNNNTLIEEMLYLIIMLGERFNPQGVQV 772
QY 768 TKEBVTWREIITHLILCIEPMPSATAKNLPENENNETGLENVINKVATFKKPGVSGHGYE 827
Db 773 AATDEIKRELIHQLSIKPMASSELVKSLPDENKETGMEVSIEVAHFKKPLGTGRGME 832
QY 828 LKDESLKDFNMVYHYHYSKTQSHKAEHMOKKRRKQENKDEALPPPPPPPCPAFSKVINLL 887
Db 833 LKPECAKEFNLYFYHFSRAEQSKAEQAQRKLRKENKEDTALPPALPPCPLFASLVNLL 892
QY 888 NCDIMYILRTVFERAIDTDSNLTWTEGMLQMAFHIALGLLEKQOLQKAPBEEV-TFDF 946
Db 893 QCDVMLYIMGTILQWAVEHHGSAWSESLQRYLHLIGMALQEKHEHLENAVEGHVQTF 952
QY 947 YHKASRLGSSAMN---IQMLLEKLGIPOLGOKDMITWILQMFVTVKRLREKSKLIVAT 1003
Db 953 TQKISKPGDAPHNPSILAMLETQNASPLSAHKADMIIRWLLKMFNAIKKIRE--CSSSP 1010
QY 1004 TSGSESIKNDEITHDKEKAERKAEARLRQKIMQASALQKNFIETHKLMYDNTSEM 1063
Db 1011 VAEAGTIMESSRDQKAEKRKAERKAEIARLRREKIMQASEMORHFIIDENKELFOOTLEL 1070
QY 1064 PGKEDSIMEESTPAVSDYSRIALGPKRGPSVTEKEVLTICLCBEQEVYKIENNAMVLSA 1123
Db 1071 DTSASATL--DSSPPVSDAALTALGPAQTQVPEPRQFVTCILCBEQEVTVGSRAMVLA 1128
QY 1124 CVQKSTALTQHRGPIELSGEALDPLFMDPOLAYGTYTSGCHVHVHVCWKYFEAVQLS 1183
Db 1129 FVQRSTVLSKDRTKTI-ADPEKYDPLFMHPDLSCGTHTSGCHGVHVAHCWQRYFSVQAK 1187
QY 1184 SQQ-----RIHVDLFDLESGEVLCPLCKSLCNTVPIPILOPKINSENADALQTLTA 1238
Db 1188 EQRQQRRLHTS-YDVENGFEFLPCLECLSNVTIPLL-LPPRSILSRRLN-FSQOPDLA 1244
QY 1239 RWIOTVLARISGYNI---RHAKGEN-----PIPIFFNMGMDSTLEFHSILSF 1283
Db 1245 QWTRAVTQQIKVVQMLRRKKNAAADTSSSESTEAMNIIPPEGRFP-----DFY----- 1292
QY 1284 GVSESIKYSNIKEMVILFATTIYRIGLKVPPDERDPVPMLTWTSTCAFTIOAIBNLJGD 1343
Db 1293 ---PRNPYSDSIKEMLTFTGTAAYKVLKHPNEGDPVPLCLMGTCAVYTIQSIERILSD 1349
QY 1344 EGRPLFCALQNRQHNGLKALMQFAVAQRITCPQVLIQKHLVRLLSVLPNKSIEDTPCLL 1403
Db 1350 EEKPVFGPLRCRLDDCLRSILTRFAAAHWTVALLPVVOGHFKFLASLPSDSYEDLPCIL 1409
```



```
QY 1404 SIDFLVHLVAVLAPSLYWDPPVDLPQSSVSSYNHLYLFLHLITMAHMLQIILL--TVD 1460
D 1410 DIDFHLVGLVLAFFALQCOD-----FSGSSLATGDLHFLHLVMAHIVOLLTSCTEE 1464
QY 1461 TGLPLAQVQEDSEEAHSASSFFAISOYTSIGSICDIPGWYLVSLKNGITPYLRCAALF 1520
D 1465 NGMD--QENPTGEBELAILSLHKLTHOYTGSALEAPSGWHLWSVRAAIMPFLKCSALF 1522
QY 1521 FHYLLGVTPEELHTNSAEGEYSALCSYLSLPTNLFLLFQBYWYVTRPLRQRCADPALL 1580
D 1523 FHYLVNGVPAPPDQV--SGTSHFELCNLYSLPTNLHLFQBNSDIMNSLIESWCQNSVK 1581
QY 1581 NCLKQKVTYVYPRKRNLSLIEPLDDYSCLLNQASHFRCPRSADDERKHPVLCFCGAILC 1640
D 1582 RYLVNGERGAISYPRGANKLIDLDPEDYSLNQASHNFCSPKSGGDKSRAPTCLICVCGSLC 1641
QY 1641 SONTCCQEIYNGEVEGACIFALHC-----KARGCAYPAPYLDXY 1680
D 1642 QSVCQAELGEDEVGACTAHTYSCGSGAGIFLVRRECOVLFLAGTKGCGYSPPYLDY 1701
QY 1681 GETDPGLKRNPLHLSRERYKRLHLVWQHCIIIEIARSOETNOMLFGFNQML 1734
D 1702 GETDGLRRGNPLHLCQERFRKIKLQWQHSITEIGHAQEAQNTLVGIDWQHL 1755

RESULT 8
O15057 PRELIMINARY; PRT; 1275 AA.
AC O15057;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein KIAA0349 (Fragment).
GN KIAA0349.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=97349984; PubMed=9205841;
RA Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 4:141-150(1997).
DR EMBL; AB002347; BAA20806.1; -
DR InterPro; IPR000408; Reg_chromatins.
DR SMART; SM00184; Znf_ring.
DR SMART; SM00184; RING; 1.
DR PROSITE; PS00626; RCCL_2; 1.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 1275 AA; 101FF1F66E056066 CRC64;

Query Match 33.5%; Score 3062; DB 4; Length 1275;
Best Local Similarity 46.4%; Pred. No. 1.3e-204;
Matches 600; Conservative 242; Mismatches 391; Indels 60; Gaps 20;

QY 484 LKYLISLPTTWRLRMOLEGRFSLKILTCMQMGEIRROVGOHIEVDPDWEAAIAI 543
D 1 LKYLISLPTWSEDELKQKFLLEGDFDAFLLELKCQMGMDPTIRQVGOHLEMEPEWAAFTL 60
QY 544 QMOLKNILMFQECACDEELLYAYKECHKAVMRCSTSFSSSKTVVQS--CGHSLETKS 602
D 61 QMKLTHVSMQWDCASDEKVLIEAYKKLAVLMQCHGYTDGEQPTILSICGHSVEIR 120
QY 603 YRVSEDLVSHLPSRLTAGLHVLRLGAVSRHLHEFVSFDFQVEVLVEPLRLCLVLA 662
D 121 YCVSQEKVSHLPSVRLLAGLHLLSKSEVAIKFPPELLPLSELSPMLLIEHPLRLCLVLA 180
```

```
QY 663 QVVAEMWRNGLSLISQVYFYQDVKCREEMDKDIIMLQIGASLMDPNKFLLLVLRQVEL 722
D 181 QVHAGWRRNGSLVQNIYYIHNVKRCRREMFQDKVVMQLQGTGVMMDPNHFLIMLSRREL 240
QY 723 AEAFN-----KTISTK--DQDLIKQNTLIEMLQVLIIVIGERYVPGVNTKEBVTMR 775
D 241 YQIFSTPDYGRFSSEITHKDVVQNNLTIEEMLYIIMLVGERSPGQGVNATDEIR 300
QY 776 EIIHLLCTEPMPHSAIAKNPENNETGLEVINVKATFKKPGVSGHGVYELKDESLDK 835
D 301 EIIHOLSIPKMAHSELVKSLPEDENKETGMESVIEAVAHEKPKPLGTGRGMYELKPECAKE 360
QY 836 FNMYYHYHYSKTOHSAEHMOKRRKQENKDEALPPPPPPFCFAPAFSKVINLLNCIDIMYI 895
D 361 FNLHYFHSRAEQSAEEAQRKLKQNRREDTALPPVLPFCPLFASLIVNLIQSDVMLCI 420
QY 896 LRTVFERAIDTNSLWTEGMLQMAFHILALGLEEKQOLQKAPEE--VTFDFVHKASRLG 954
D 421 MGTILQWAVEHNGYAWSESMLQRLVHLIGMALQOEKQHLNTEEHVVTFTTQKLSKPG 480
QY 955 SSAMN---IQMLLEKLGIPQLEGQKDMITWILQMFDTVKRLREKS--CLIVATTSGESI 1010
D 481 EAPKNSPSILAMLETQNAPLYEVHKDMIRWILKTFNAVKKRESPTSPVAETECTIM- 539
QY 1011 KNDEITHDKEAERKKAARLHROKIMAOASALQKNFIETHKLMYDNTSEMPGKEDSI 1070
D 540 --EESRDQKAEKRRKAEIARLRREKIMAOSEMQRHFIDENKELFOOTLELDASTAV 597
QY 1071 MEEESTPAVDYSRIALGPKRGPSYTEKEVLTCILCOEEOEYKIENNAVLSACVOKSTA 1130
D 598 LDH--SPVASDMTLTALGPTQVPEQOFVTCILCOEEOEYKVESRAMVLAIFYORSTV 655
QY 1131 LTOHRGKPIELSGEALDPLFMDPDLAYGTYTSCGHVMHACVQKYEAVQLSSQO--- 1186
D 656 LSKNRSKFIQ--DPEKYDPLFMPHDLSCGTHTSSCGHIMAHACWQRFDSVQAEQRQR 714
QY 1187 -RIHVDLFDLESGEYLCPLCKSLCNTVPIIPLQPKINSENADALQALLTLARWTQTVL 1245
D 715 LRLHTS--YDVENGFEFLCPLCECLSNVTIPL-LPPRNIFNNRLN-FSDQPNLTQWLTIS 771
QY 1246 ARISQYNIRHAKGNPIPIFFNQMGCDSTLEPHSLISFGVSESIKSNSEKENVILPAT 1305
D 772 QOIKALQFLRKEESTPNNASTKNSENVDELQLEPEGRFPRPKIPYSESIKEMLTFTFTA 831
QY 1306 IYRIGLKVPPDRPRVPMLTWSTCAFTQATENLGLDGEKPLFGALQNRHNGKALMQ 1365
D 832 TYKVLKLVHPNEEDPRVPMCGSCAYTIQSIERILSDEDEKPLGFLPCLDCLSLTR 891
QY 1366 FAVAQRITCPQVLIQKHLVRLSLVLPNIKSEDTPCLLSIDLPHVLVGAVALFAPSLYWD 1425
D 892 FAAAAHTVASVSVQGHFCKLFASLVPNDSHEELPCILDIMFHLVGLVLAFFALQCOD 951
QY 1426 PVDLPSSVSSYNHLYLFLHLITMAHMLQIILL--TVDTGLPLAQVQEDSEEAHSASSFF 1482
D 952 -----FSGISLGTGDLHFLHVTMAHIIQILLTSCTEENGMD--QENPCEESAVLALY 1004
QY 1483 AEISOYTSIGSICDIP--GWLAVSLKNGITPYLRCAALFHYLLGVTPEELHTNSAEGE 1541
D 1005 KTLHOYT--GSALEKIPESGWHLWRSVRAGIMPLFKCSALFHYLVVSPDPDIQV--PGTSH 1062
QY 1542 YSALCSYLSLPTNLFLLFOEYWDVTRPLRQRCADPALLNCLKOKNTVYVYPRKRNLSIE 1601
D 1063 FEHLCSYLSLPLNNLCLFQENSEIMNSLIESWCNRSNVEKRYLSEGRDAIRYPRESNKLIN 1122
QY 1602 LPDDYSCLLNQASHFRCPRSADDERKHPVLCFCGAILCSQNTCCOEVNGEVEGACIFH 1661
D 1123 LPEDYSSLLNQASHNFCSPKSGGDKSRAPTCLICVCGSLCSQSYCCOTELGEDVGACTAH 1182
QY 1662 ALHC-----KARGCAYPAPYLDYVETDPTGLKRGPNLHLSRERY 1701
D 1183 TVCSGSGVGIPLVRRECOVLFLAGTKGCGYSPPYLDYVETDPTGLKRGPNLHLSRERY 1242
QY 1702 KLHLVWQHCIIIEIARSOETNOMLFGFNQML 1734
```


[illegible]

```

539 EVTHKDVQONNTLIEEMLYLIMLVCERENPGVGVAATDEIKREIIHQLSKPMHSE 598
791 IAKNLPENNETGLENVINKVATFKPGVSGHGVYKDESLKDFNMVYFHYSKTHSK 850
599 LVKSLPEDENKETSIVESAHEFKKPGLTGRGMVELKPECAKEFNLYFHSRAQSK 658
851 AEBHMKRRKQENKDALPPPPPCPAFSKVINLLNCDIMMYILRTVFERAIDTDSNL 910
659 ABEAQKLRKREKNDTALPPALPPCPFLFASVINLQCDVMYIMGTILQWAVEHHGSA 718
911 WTEGMLQAFHIALGLLEKQOLQAPPEEV--TDFYHKASRLGSSAMN--IQMLEK 966
719 WESMLQRLVHLGLMALQEKHLENVAGHVQTFQTKISPGDAPHNSPILAMLET 778
967 LRGIPLEGOKDMITWILQDFVTKRLREKSCULIVATTSGSESIKNDIETHDKERKR 1026
779 LQAPSLEAKHOMIRLLKMFNAIKIRE--CSSSPVAEAGTMEESSRDKKAERKR 836
1027 KAEARLHRQKMAQMSALQKNPIE 1051
837 KAEIARLRREKIMAOEMFEMORHFI 861

RESULT 12
Q8SX71
ID PRELIMINARY; PRT; 1824 AA.
AC Q8SX71;
DT 01-JUN-2002 (trEMBLrel. 21, Created)
DT 01-JUN-2002 (trEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (trEMBLrel. 23, Last annotation update)
DE LD31957p.
GN CG9086.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RC SEQUENCE FROM N.A.
RA STRAIN=Berkely.
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celiker S.;
RA Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AY094815; AAM1168.1; -.
DR FlyBase; FBgn030809; CG9086.
DR InterPro; IPR003126; znf_Nrecognin.
DR InterPro; IPR001841; znf_ring.
DR Pfam; PF02207; zf-UBR1; 1.
DR SMART; SM00184; RING; 1.
DR SMART; SM00396; znf_UBR1; 1.
SQ SEQUENCE 1824 AA; 208329 MW; D68FCB81D35C7124 CRC64;

Query Match
Best Local Similarity 24.2%; Score 2210.5; DB 5; Length 1824;
Matches 596; Conservative 300; Mismatches 700; Indels 323; Gaps 56;

QY 10 EAGGTERMEISALPOTPORLASWDDQVDFTAFHLHLAQVPEIYFAEMDPDLKQEE 69
DB 29 QAGTLDLRSDIIFELKRESKPR---YFDYQT---SATVKDNTVITLKCMEKE---SLAKEE- 78
QY 70 SVQMSITFTPLEWLFGEEDPICLEKLHSG-AFOLCGRVFKSGETYSRCDAIDPTCVL 128
DB 79 -----IIDVVVEMLGDNPSALEKRLLEGNTATVCGKFKNGEPTYSRCGVDPTCVL 133
QY 129 CMDCFODSVHKHRYKMHMTSTGGGFCDCGDTBAWKTGPFCVNHPEGRAGTIKENSRCPLN 188
DB 134 CVNCFKRSARHFKYKMTSGGGGCCDCGDEAWKKDQYCELHLANKRNPLESKI---LT 190
```

```

189 EEVIVQARKIFPSVIKYVEMTIWEEKELPPELOIREKN-----ERYVCVLPNDEH 240
191 DAVLERVEICGAILAFCSYL-----ETEPNASLQCLDGNVEGGQVDGAQYCTVYNDES 246
241 HSYDHVYISLQALDCELAEAQLHTTADKEGRRAVKAGAYAAQCAQEKEDIKSH----- 294
247 HTFQVQITQTKIAKCRAKDAWEIVAAIDREGRAVVKCDTFECNKLKVKYSIENQMLPTIS 306
295 --SENVSQHPHVEVLHSEIIMAHQKFAIRLGLSWMNKMISYSDFFRQIFQACLRPEPDS 352
307 LVSTARNQSLRTSVLHIGAVACQAFQALQLGWFQEFVHRVHLFRKTFSELVQKQ---E 363
353 NPCLISRLMLWDKILKARKLHLEIFSSPFMEMEYKLFAMEFVKYKQLOKEVISDD 412
364 TFC-IRHILEYDVWLKWTARTCWRLHLLISGMLMDNKMILAQEFSRRYATIVEDISD 422
413 HDRSISITALSVOQMTPTLARHLLLEEQNVISVITETLLEVLPEYLDNRNKNFQF--YS 470
423 HDHAFSIVLSVQVLTVPVPSIAHHLHAHEGIFDKLLHTFYHVAIEKFINRKTLLHFSKNIAS 482
471 QDKLGRVAVICDLKYILISKPTIWTRELRLMQFLEGFRSFLKILTCMQGMEELRROVGQH 530
483 LTFEKRANYILYDLRYLLSLKAPDVLNDRNGFLEGCRALMRVLNVMQGMESMTROTGOH 542
531 IEVDPDWEAATAIQMLKNIILMFQEWACDEELLVAYKECHKAVMRGCTSFISSTKTV 590
543 MDYEPWECAFNHLKATTTISQVIDWASGDVKLLKLYKMTMRALV--SNSFIVGGEKV 600
591 VQS---CGHSETKSYRVSEDLVSIHLPLSRTLAGLHVRLSLGASVRLHEFVSDFQV 647
601 MQPKVADHVANCLVYDISQVPSIHLPLSRFFAGIYL---HLGA---HD-LTYDGLQT 652
648 EV-----LVEYPLRCLVLVAQVVAEMRRNGLSLSQVYVYODVKCREMYDKD 696
653 ETEALSTKLPRELIEPVLCQAMIAQVAGLWRRNGYTLHLQLFYRNVRCRMLDRD 712
697 IIMLQIGASLMDPNKFLLLVLQRYE---LAEAFNKTIISTKD---QDLKQYNTLIEMLQ 750
713 IACIQIGASLMSNEFLIHLNRRNTIPWLQENYWSLLSGNEMNDIIRE-ASIFDEPLE 771
751 VLIYIVGERYPVGVNVTKEVMTREITHLICIEPMPHSAIAKNLPENB--NNEGTLENV 808
772 LLVIIVIGERMPGVMVTEEDRLKEITLQCLIKPYSHELSEALPDGNSGNSDNVFEV 831
809 INKVATFKKP-GVSGHGVYELKDESLKDFNMVYFHYSKTHSKAHMKRKRQENKDEA 867
832 INTVAFKKPVGADSKGVYELKEHLKFEFMFYHTKDEKSAELQERKRAKKQLVC 891
868 LPPPPPPFCFPAFSKVINLLNCDIMMYILRTVFERAIDTDSNLWTGLMQAFHIALGL 927
892 CPPPMLPKLTPTAFTPMANILOCPVFLNCSLIMERALNAYSFSFTESHLQKVLHLLGYAI 951
928 LEEKQLOKAPEEVTFDFYHKASRLGSSANNIOMLEKLGIPOLEGOKDMITWILQMF 987
952 QEE-----LSEHYPPFLSFYERSOYFG-----ILEKLEELARCPLEAHYDVLVTIERF 1000
988 DTVKRLREK-----SCLIVATTSGSESIKNDIETHDKERKAEKRAARLHRQK 1037
1001 ---KOLQAKQAPSDGRAPSC-SOQGTGKGLSLSAE---QAREENRRLAERAAH 1052
1038 IMAQMSALQKNFIETHKLMDYNTSEMPGKE-----DSIMEESTPAVSDYSRIA-L 1087
1053 IMAQMQAQAQSFISNAEMFADTNETRKESASTGPMDWEDIPPEEQGAVALESKVACL 1112
1088 GPKRGPSTKEVLTICQEEQEVKIENNAVLISACVQKSTALTQHRGKPIELSGEALD 1147
1113 GPKRKFYHGTDDTFKILCFENCALISRGRLVSSAFVQTSRV----- 1155
1148 PLFMDPDL---AYGTGSCGHVMAVCQYKF---EAVQLSSQRIHVLDLDESSEVL 1201
1156 -IETPTPLNQSQSAHISCCGHVMHYSWLEFYFTNEEFKELRRHRNRAALAAQAAVNEFQ 1214
1202 CPLCKSLCNTVPIPIIPLQPP-----KINSENADALAQL-----LTAR 1239
```

```
Db 1215 CPCRTLSNIIPTVETLPAFAAPPSPNESYLPDSEVMSLAIELGNVKBHETLTP 1274
QY 1240 WICVTILARISGYNIRHAKGNPIPIFNQMGDSTLEFHSILSGFVSESSIKYSIKEMV 1299
Db 1275 SVSNIL-RLSGVVGGLAQPERSVOLIKNPP-----RLHADIETGIE-----FLKKA 1319
QY 1300 ILFATYIYRIGLK-VPPDERDRVPMLTWTCAFTTQAIENLGLDDECKPLFGALONQHN 1358
Db 1320 LNTMKIQSHLKHDPFAIESIEMVPII-WDSCSYTTLQALETYIYAVEKPLKAELSMRHQ 1378
QY 1359 GLKALMO-----FAVAQRITCPVOLIKHLVRLSVVLPNIKSDTCLLSIDLFHVLV- 1412
Db 1379 CARNLVACRSRSALEWETDLPLPMPRSQAERSSRLDIFNDONTSVLEWDCFRVLVP 1438
QY 1413 ---GAV-LAPPSLYWDDPVLQPSVSSSYNHLILFHLITMAHMLQILLVDTGLPLAQ- 1467
Db 1439 FQGVNLMLVPEKGYKTIIP-----SGSMFDVFMOTMFLAQLTKAVLCFVDEKAKR 1492
QY 1468 -----VOEDSEE--AHS 1477
Db 1493 AEKAPNELTQLDYIEQLPSRIRDNMTDFYRRYNIPARVLQTKQKQLVEESEENQGH 1552
QY 1478 -----ASFEEAISQTSSTGCCDIPGWYLVWSLKGITPYLRCALFPHYLGLGTPP 1530
Db 1553 QTVIPCESHLLALLEY-----VORQSSFLRCSCLFRFLTDVDFP 1595
QY 1531 EELHTNSAEGEYSALCSYLSLPTNLFLLFOEYWDVTRPLL-----QRRCAD 1576
Db 1596 DTEPTDQPD-RFDLMQCYGLDPLMGVYFD--METVYATMHSFASHPHIDREVEQRCP 1652
QY 1577 PA-----LLNCLQKQNTVYPRKRNLSIELPDDYSCLLQNAQHFRCPSRADDKHPVL 1631
Db 1653 DARRSLQVWPCLRP-----LPR-----LKVLCDDFSILNSVSDIFCPNNEREEMKPTM 1702
QY 1632 CLFCGATLCSQTCQCEVNGEEVGCIFHALHC-----KARGCAY 1672
Db 1703 CLICGLTLCQCYCCQPELQKSVGATHAHAGAEVGLFTRDCQVYVLRGKGCYFV 1762
QY 1673 PAPYLDEYGETDGLKRGNPLHLSRERYRKLHLWQOHCITEELARSQETNOMLFGFNW 1731
Db 1763 PPPYLDEYGETDGLRGNPLRLSQAARYKIYQLWGLHGLGETARLNDANVANVAQAQW 1821

RESULT 13
Q9VX91 PRELIMINARY; PRT; 1824 AA.
AC Q9VX91:
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE CG9086 protein.
GN CG9086.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
```

```
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Foubler K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasok P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson S.M., Nelson K., Nixon K., Nusskern D.R., Pacbe J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacbe J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Shouanenavong S., Strong R., Svirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RA "Sequencing of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RA "Annotation of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003504; AAF48687.2;
DR FlyBase; FBgn0030809; CG9086.
DR InterPro; IPR003126; Znf_NrecoGn.
DR Pfam; PF02207; zf-UBR1; 1.
DR SMART; SM00396; ZNF-UBR1; 1.
SQ SEQUENCE 1824 AA; 208359 MW; 79992c11175e82a0 CRC64;

Query Match 24.2%; Score 2208.5; DB 5; Length 1824;
Best Local Similarity 31.1%; Pred. No. 8.1e-145;
Matches 595; Conservative 300; Mismatches 706; Indels 313; Gaps 55;
```



```

RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium. ";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Gattung S.;
RT "The sequence of C. elegans cosmid C32E8. ";
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RT "Direct Submission. ";
DR EMBL; U88308; ABA42328.1;
DR WormPep; C32E8.11; CE08535.
DR InterPro; IPR003126; Znf_Nrecognin.
DR Pfam; PF02207; zf-UBR1; 1.
DR SMART; SM00184; RING; 1.
DR SMART; SM00396; Znf-UBR1; 1.
KW Hypothetical protein.
SQ SEQUENCE 1927 AA; 218097 MW; 6C2A581C52A8E9CF CRC64;

Query Match 20.68; Score 1886.5; DB 5; Length 1927;
Best Local Similarity 27.38; Pred. No. 2.6e-122;
Matches 553; Conservative 305; Mismatches 636; Indels 513; Gaps 67;

QY 102 QLAGRVFKSGETYSRCDAIDPTCVLCMDFQDSVHKHRYKMHSTGTGGFCDCGDTGA 161
DB 14 QICGHVEKNGELYTCIDCATDCTCVMLCQCFEYSIHSHKXKMHSSGSGCYCDGDADA 73

QY 162 WKTGPPFCVNNHPG-----RAGTIKENSRCPLNEEVIVQARKTFFSPVYKVVEMTWE 214
DB 74 WTEGYACANHEKKDEEAVALPELKK--RCEQLVEIILQ---FSLSMITHKDKLKLPEI 128

QY 215 EKELPPELQIREKERYCYVLFNDEHSYDHVYLSQALDCELAELAQHTAIDKEGR 274
DB 129 FEKKPKE--VTNEAQQQVTLVYNDETHYTSVIVKLYTHCTKDOAMLVATIVDREGS 186

QY 275 AVKAGAAQCEAKEDIKSHS-----ENVSOH--PLHVEVLHSETMAHOKFALRLGSW 325
DB 187 AVLKGSXADCTKAKDDVQRTADPTSIKSSNNHPLSVKVMYDTTFLALQNFSLTLTW 246

QY 326 MNKIMSYSSDFRQIFCOACL----- 345
DB 247 LNTQMDVFPPLREIVGEILLSSKFAKKNYTRKMKSEDRQLVAGIIRNVMLPDDEEEL 306

QY 346 -----REEP-----DSENPLCI 357
DB 307 FALDGRMDVEMDDDDDIGEALQMEIDADDEEITAAALAGVSEHQSPGSRDSSSTETML 366

QY 358 SRLMLDAKLYKGARKTLHELIFESSFFMEYKKLFAEFVKYKQLQKEYISDDHORSI 417
DB 367 ENILLQDTQMKAGRSILHOMLMTVMFYDQKVRFAKAFMLHYNEIYEDFIKDDHEMDV 426

QY 418 SITALSVMQMTVPTLARHLIEEQNVISVITETLLEVLPEYLDNRN-----KFNFGQYS-Q 471
DB 427 SVVGLSVQFMTVPSLARKLVAEDQAFSVISKAIRDQTDKFKVYVYNDGKIARFDTSRFP 486

QY 472 DKLGVRVAVICDLKYILISKP--TIWTERLMQFLEGFRSPKLTLCMQGMEETRRQVGQ 529
DB 487 PELRRSLHTRDMAYILNAPVSESDWNRELIDGFGVQGFADFLFLQHLQGMDEVKQRAVE 546

QY 530 HIEVDPDWEAIAIQMLKNILLMFQWCACDEEL---LLVAYKECHKAVMRCSTSTFISS 586
DB 547 HQWSESEWETAFTNLLRLKDAISMIIIGWAETNEEVHRLMTM---CLELMNRMPVPTKS 603

QY 587 S-----KTVQSCGHSLETKSYRVSEDLVSTHLPLSRTLGLAG-----HV 625

```

```

DB 604 EEDTYELTVTINGESCRISHFDVLKASSTSVHQPVVRIITAGLFSASNYTGFLNRRSNHSHT 663
QY 626 RLSRLGAVSRHEHVFVSFEDFQVEVLVEPLRCLVLVAQVAVMWRNRGSLISQVFPYQD 685
DB 664 SLNQ-----ERIKELINCKD--ETNLYELSLRVLVLAQSNATLWRRNGFSLNIQHNYFS 717
QY 686 VKCREEMVDKDIIMLQIGASLMDPNKFLLLVLQRYEL-----AAFPNKITSTKQD 735
DB 718 PLCRNEMFDRDVLMMQVGAALTPTKTFIHLHLRFLFKWATSEFQDQKANEKPAKPESE 777
QY 736 DLIKQYNLIEEMLOVLIVIGERYVPGVGNVTKEEVTMRIIHLILCIEPMHPHSAKAKNL 795
DB 778 DLSKTLVMAEEFTQCLILILICERTYGVGKTRPMDQMKAREVIHILICTGSHTSFHIQQKM 837
QY 796 PENENNE--TGLENVINKVATFKPGVSGHYVELKDESILKDFNMFYHYHQSHKAEHM 854
DB 838 SHDINSKRLSLHEAVNLVADFRKPLATTAGOFCHKESSLPTYSPEFPMHYKSDQSAEQS 897
QY 855 QKK--RRKOENKDEALPPPPPEFCFAPSKVINLLNCDIMMYILRTVFERAIDTSDNLWTE 913
DB 898 QARVRAKMEKSVRACAPPILPDFQTFEPIPELTITNLIHVVRLLIDRTA--RRSRFSSD 956
QY 914 GMLQMAFHILALGLEEKQOLQKAPEEVTFDYHKASRLGSSAMNIOMLLEKLGIPOL 973
DB 957 RLFHKTLYLIGIALNEE-----EKNPSPGFTQRA---EESIGLLSLLEGLVGKP-- 1002
QY 974 EGQKDMITWILQMFDTVKRLREKSLIVATTSGSES--IKNDEITHDKAEKRAKRAEA 1031
DB 1003 --ESSICIPILLEV--TVEKYRK--LIKARAGVPEAAPAPENKPAQSEIEIKAKRAARA 1055
QY 1032 RLHRQKIMAQSAQKQNFETHKLMYD-----NTSEMPG-----KED-----STMEERS 1075
DB 1056 EM-RQKAMAKMSNMQSKEM--KKTEDEBKDESQTPSEKSESVYVKDDYDNKHFDEVD 1111
QY 1076 TPVSDYDRIALGP--KRGPSVTEKEVILFCILQEQEV--KIENNAWVLSACVQKSTALRQ 1133
DB 1112 VKQVGHDPVPCIGANKWHAELVKPRTLTLCILOEDEIIPAOQKPMVCVCAAFIQQSOLF-- 1170
QY 1134 HRGKPIEL-----SG--EALDPLFMDPLAYGTYTGTSCGHVMHVAWCWKYFEAVOLSSQOR 1187
DB 1171 HKNKGELMTASSGISTRDLLTAPATQYGVVDVSTCSHMYEYCTRSIAENRSRESUR 1230
QY 1188 IHV-----DLFDLESGEVLCPLCKSLCNTVPIIP--LOPK-----INSENADAL 1231
DB 1231 ARQVQCHSHKWDVTENGEOCPCLCKRLSNAAIPVLPAYOLTNQNGFSTVSGAKENFDT- 1289
QY 1232 AOLITLARKIQTVLARI-----SGYNIRHAKGENPI-----PIFFNQGMG 1271
DB 1290 -----WVARVKRNLEMPLESSESVSKGHSKRSHRSERSLDLLEKLSKDPDTANTSAG 1341
QY 1272 DSTLEFHSILSGV-----ESSIKYSNSIKEMVI----- 1300
DB 1342 -----VLOGGAMEMSSATHMPASBSOMLMTTSPSODDVEFYNELAAMFVDQVNNVT 1393
QY 1301 ---LFATTIVRIG--LKVPDPDERPVRPM----- 1324
DB 1394 TSPAATPETIPAIGSSSRIPESQESGKKPLSQIOHVLVSLIRPPPALINSNRCSSPE 1453
QY 1325 -----LTWSTCAFTIOAIENLGLDEG 1345
DB 1454 GFEEPIKDLGNMKMFRKRGNELKTNFIEKHLKGVYISTVWQSTAHVARAISTSYLHYDK 1513
QY 1346 KPLFCALQNRQUNGKAL-----MQFAVAQRITCPQVLIQKHLVYR-LLSVVL--- 1391
DB 1514 KPLFCALATNRQDRCLSAMARLCAISLHNMQFL-----HAYSDMLRVLFLCEP 1560
QY 1392 PNKISEDTPCLLSIDLHVLVGAFLAPSLYWDPDVLOPSSV--SSSYNHLVLIHLITWA 1450
DB 1561 PRPKLAQTP-----GSPLLSAPSTSSFTPA---PAQIPHSGTFAFVLQVLPNPA 1606
QY 1451 -----HMLQILLTVDITGLPLA-----QVQEDSESAHSASF 1481
DB 1607 GPRKNVNLILQIDI-----LSLIAITHSEADGNDVNMEEBQESQMEVDPVAAQIRKL 1661

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 1, 2002, 12:39:35 ; Search time 4409.56 Seconds
(without alignments)
23599.681 Million cell updates/sec

Title: US-09-724-126A-1
Perfect score: 6308
Sequence: 1 gccaaagattcgacgagg.....aattttgtattgtgtttt 6308

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues
Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- GenEmbl: *
- 1: gb.ba.*
 - 2: gb.htg.*
 - 3: gb.in.*
 - 4: gb.om.*
 - 5: gb.ov.*
 - 6: gb.pat.*
 - 7: gb.ph.*
 - 8: gb.pl.*
 - 9: gb.pr.*
 - 10: gb.ro.*
 - 11: gb.sts.*
 - 12: gb.sy.*
 - 13: gb.un.*
 - 14: gb.vi.*
 - 15: em.ba.*
 - 16: em.fun.*
 - 17: em.hum.*
 - 18: em.in.*
 - 19: em.om.*
 - 20: em.or.*
 - 21: em.ov.*
 - 22: em.pat.*
 - 23: em.ph.*
 - 24: em.pl.*
 - 25: em.ro.*
 - 26: em.sts.*
 - 27: em.sy.*
 - 28: em.un.*
 - 29: em.vi.*
 - 30: em.htgo_hum.*
 - 31: em.htgo_inv.*
 - 32: em.htgo_rod.*
 - 33: em.htgo_hum.*
 - 34: em.htg_inv.*
 - 35: em.htg_rod.*
 - 36: em.htg_other.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

*

Result No.	Score	Query Match	Length	DB	ID	Description
1	4435.4	70.3	6395	6	AR030784	AR030784 Sequence
2	4435.4	70.3	6395	6	AR121463	AR121463 Sequence
3	4435.4	70.3	6395	10	AF061555	AF061555 Mus muscu
4	2546.8	40.4	2550	9	AK027803	AK027803 Homo sapi
5	996.2	15.8	1001	6	AR030785	AR030785 Sequence
6	996.2	15.8	1001	6	AR121464	AR121464 Sequence
7	994.2	15.8	999	9	AF061556	AF061556 Homo sapi
8	981	15.6	6158	9	AB002347	AB002347 Human mRN
9	764.2	12.1	818	9	HSX521	247040 Human partI
10	651.6	10.3	2958	9	AK026998	AK026998 Homo sapi
c 11	503.8	8.0	166518	9	AC090514	AC090514 Homo sapi
c 12	487.8	7.7	2512	9	AK026948	AK026948 Homo sapi
c 13	354	5.6	190727	9	AC068724	AC068724 Homo sapi
c 14	279.4	4.4	61896	2	AC090992	AC090992 Homo sapi
c 15	215.2	3.4	910	10	MM1UBR2	AF067372 Mus muscu
c 16	212.2	3.4	107304	2	AC016274	AC016274 Homo sapi
c 17	212.2	3.4	164468	2	AC021899	AC021899 Homo sapi
c 18	180.8	2.9	164468	2	AC021899	AC021899 Homo sapi
c 19	160.6	2.5	3980	9	HS1UBR4	AF067383 Homo sapi
c 20	150.2	2.4	107304	2	AC016274	AC016274 Homo sapi
c 21	132	2.1	28834	2	AC017982	AC017982 Drosophil
c 22	132	2.1	152545	3	AC010922	AC010922 Drosophil
c 23	132	2.1	179016	3	AC018489	AC018489 Drosophil
c 24	130.4	2.1	900	9	HS1UBR3	AF067382 Homo sapi
c 25	123.4	2.0	907	10	MM1UBR1	AF067371 Mus muscu
c 26	115.2	1.8	61896	2	AC090992	AC090992 Homo sapi
c 27	111.8	1.8	2616	10	MM1UBR6	AF067376 Mus muscu
c 28	106.4	1.7	830	9	HS1UBR5	AF067384 Homo sapi
c 29	104	1.6	143291	9	AL137792	AL137792 Human DNA
c 30	104	1.6	143697	2	AL358572	AL358572 Homo sapi
c 31	104	1.6	295265	2	AL353354	AL353354 Homo sapi
c 32	102.6	1.6	716	10	MM1UBR5	AF067375 Mus muscu
c 33	101.8	1.6	1000	9	HS1UBR6	AF067385 Homo sapi
c 34	100.8	1.6	1370	9	HUMORF02	D14658 Human mRNA
c 35	100.4	1.6	1100	10	MM1UBR4	AF067374 Mus muscu
c 36	97	1.5	179588	2	AC009852	AC009852 Homo sapi
c 37	97	1.5	182660	2	AC012416	AC012416 Homo sapi
c 38	97	1.5	183982	2	AC009825	AC009825 Homo sapi
c 39	91.4	1.4	1000	9	HS1UBR7	AF067386 Homo sapi
c 40	91.2	1.4	727	10	MM1UBR7	AF067377 Mus muscu
c 41	89	1.4	789	4	CFU12687	U12687 Canis famil
c 42	88	1.4	168639	2	AP001992	AP001992 Homo sapi
c 43	88	1.4	213613	2	AC013405	AC013405 Homo sapi
c 44	86.4	1.4	400	9	HS1UBR1	AF067380 Homo sapi
c 45	86	1.4	195064	2	AL138921	AL138921 Homo sapi

ALIGNMENTS

RESULT 1
LOCUS AR030784 6395 bp DNA
DEFINITION Sequence 1 from patent US 5861312.
ACCESSION AR030784
VERSION AR030784.1 GI:5943998
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 6395)
AUTHORS Varshavsky,A. and Kwon,Y.Tae.
TITLE Nucleic acid encoding mammalian UBRL
JOURNAL Patent: US 5861312-A 1 19-JAN-1999;
FEATURES Location/Qualifiers
source 1..6395 /organism="unknown"
BASE COUNT 1802 a 1376 c 1519 g 1698 t
ORIGIN

29-SEP-1999

PAT

Query Match		70.3%;	Score 4435.4;	DB 6;	Length 6395;
Best Local Similarity		86.3%;	Pred. No. 0;		
Matches 4997;		Conservative 0;	Mismatches 721;	Indels 69;	Gaps 6;
Qy	590	ttcaggggccgtcgtaaaagtgtcgtccctgtctc-tccgaccggccacaggtttccgct	648		
Db	8	ttcaggggccgtcgtaaaagtgtcgtccctgtcgtccgtcgccggccgacacaggtttccgct	67		
Qy	649	tgcctctggccgggggtcggaactgcaggcggtcaagtttccctcagaatggcgagcaggg	708		
Db	68	agctggcgccgggggtcggaactgcggcggttcgtttcccttaagatggcgacgaag	127		
Qy	709	aggttgagggtactgagaggtatgaaatcagcgggagttatccccagaccocctcagcgtc	768		
Db	128	agatggagcggcgccgagaggtgagcgtcagcccgagcctccctcgcccgccgacgccc	187		
Qy	769	tggcatctgtgggtatcagacaggttgatttttatactgcgtttcttgatcatattggcac	828		
Db	188	cgcatcgtgggtggatcagcaagtgtgatttcttactgcgtttcttgcattcatttggcac	247		
Qy	829	aattggtccagaaatttactttgctgtaattggaccacagacttggaagcagggagaa	888		
Db	248	aattagtcaccagaaatttatTTTGTGTGAGATGGACCCAGATTGGAAAAGCAAGAGAGA	307		
Qy	889	gtgtacaaatgtcaatatctactccactggaatgggtactttatttggagaagatccagata	948		
Db	308	gtgtacagatgtcaatactactccttttggagtggtacttatTTTGGAGGATCCGGATA	367		
Qy	949	tttgcctagagaaattgaagcacagtcagatcttcaagctttgtggagggttttcaaaa	1008		
Db	368	TTTGCCTTAGAGAAATTAAACACAGTGGAGCGTTCACGTTGTGGGAAGSTTTTCAAAA	427		
Qy	1009	gtgagagacaacctattcttgacgggattgtgcaattgatccacatgtgtactctgta	1068		
Db	428	GTGGAGAAACAACATATTCCTGTGAGGATTTGCAATTTGATCCAACTGTGTGCTCTGTA	487		
Qy	1069	tggactgcttccagacagtgcttcaataaaatcatcgttacaagatgcatacttctactg	1128		
Db	488	TGGACTGCTCCAAAGTAGTGTTCATAAAAACCATCGTTACAAGATGCATCTTCTACTG	547		
Qy	1129	gaggagggtctgtgactgtgagacacagaggcatggaactggccctttttgtgtaa	1188		
Db	548	GAGGGGGCTCTGTGACTGTGTGAGACACAGAGCGTGGAAAACCTGGCCCTTTTGTGTGG	607		
Qy	1189	atcatgaacctggagagcaggtactataaagagaattcacgctgtcgttggaatgaag	1248		
Db	608	ATCAGCAGCCTTGGAGACAGGTACTACAAAGAGAGCTTACATTTGCCCATTTGAATGAAG	567		
Qy	1249	aggtaatgtccaaagccaggaaaaatttcttccagtgtataaaatatgtcgtagaatatga	1308		
Db	668	AGGTGATTGCTCAAGCCAGGAAATATCCCTTCGGTGATATAAATACATTGTAGAAATGA	727		
Qy	1309	ctatatgggaagagaaaaaactgcctcctgaaactcagataaggagaaaaatgaaa	1368		
Db	728	CTATATGGGAAGAAAGGAATTTGCCCTCTGAACTGCAGATAAGGGGAAAAAATGAAC	787		
Qy	1369	gatactattgtctcttttcaatgatgaacaccatttcattatgaaccagtcataacagcc	1428		
Db	788	GATACATATTGTCTCTTTTCAACGATGAGCACCATTTGTTATGATCATGTGATCTACAGTC	847		
Qy	1429	lacaagagactttgactgtgagctgcagagggcccaattgcataccactgcatattgaca	1488		
Db	848	TGCAGAGAGCTTAGATTGCGAGCTTGCAGAGCGACAGCTGACACAGCTGCCATCGACA	907		
Qy	1489	aagagggtcgtgggctgttaaagcgggagcttatgtcgtcttggcaggagcaaaaggaag	1548		
Db	908	AAGAGGGTCGCGGGCTGTCAAAGCAGGTGTGTATGCCACTTTGCCAGGAAGCAAGGAGG	967		
Qy	1549	atataaagatcattcaaaaaatgtctctcaacatccactcttagaagattattacact	1608		
Db	968	ATATAAAGAGTCACTCAGAGAACGCTCTCTCAGACACCCCTCCATGTGGAAGTGCTGCAC	1027		

Qy	1609	cagagattatggctcatcagaataatttgctttggctgttgggttccctgatatgaacaaaatta	1668		
Db	1028	CCGTGGTTATGCTACCAGAAATTCGCTCGCTTGGCTTGGCTATGATGAACAAATTA	1087		
Qy	1669	tgagctattcaagtacttaggcagatcctttgccaagcatgccttagagagaacctg	1728		
Db	1088	TGAGCTATTCAAGTGACTTTAGACAGATATTTTGCAGGGCTGCCTCGTAGAAGAACCTG	1147		
Qy	1729	actcgagaaatccctgtctcataaagcaggttaatgcttttggggatgcaaaactttataaag	1788		
Db	1148	GCTCTGAAAATCCCTGCTTATTAAGCAGACTAATGCTTGGGATGCAAAACTTTTATAAAG	1207		
Qy	1789	gtgcccgttaagatccctcatgaattgatcttcagcaggtttttttatggagatggaataca	1848		
Db	1208	GTGCCCCGTAAAGTCCCTTCATGAATTGATCTTTTAGTAGTTTATTTATGGAATGGAATACA	1267		
Qy	1849	aaaaactcttgcattggaatttgaaattgaaattataaaacactgcagaaagaatatatca	1908		
Db	1268	AAAACTCTTTGCTATGGAAATTTGTGAAGTATTTATAAACACTGACAGAAAGATCATATCA	1327		
Qy	1909	gtgatgatcatgacagaagtatactctataaactgcacttccagttccagatgttactgttc	1968		
Db	1328	GGACGACCCAGAGAGAAGCATCTCCATTAACGCCCTGTCCGTGCAGATGCTCACCGTCC	1387		
Qy	1969	ctactctgctcgacatcttattgaagagcagaatgttatctctgtctacttactgaactc	2028		
Db	1388	CGACCTTGGCCCCGCATCTTATTGAAGAGCAGAAATGTTATTCTGTCTATTACTGAAACGC	1447		
Qy	2029	tactagaagttttacacctgactctgacaggaacaataaattcaactctccaggttata	2088		
Db	1448	TGCTAGAAGTTTACCTGTAATACTTGGACAGGAACAATAAATTCACCTCCAGGGTTATA	1507		
Qy	2089	gccagagacaaattgggaagagtataatgcagtaataatgtgacctaaaatatactctgata	2148		
Db	1508	GCCAGGACAACTGGGAAGAGTCTACGCAGTTATATGTGACCTAAAGTATATCTTGATTA	1567		
Qy	2149	gcaaccccaaatatggacagaagaattaaagaatgcagttcccttgagggttttcgactct	2208		
Db	1568	GCAAGCTGTGCATATGGACAGAACGATTAAAGAGCGCAGTTCCCTGGAAAGTTTCCGGTCT	1627		
Qy	2209	ttttgaagattcttaacctgtatgcagggaatggaagaaatccgaagacaggttgggaac	2268		
Db	1628	TTCTGAAGATTCCTTACTGTATCAGGGAATGGAAGAAATCAGAAGACAAGTTGGACAAC	1687		
Qy	2269	acaattgaatggatcctgattgggaggtccattgctatacacagatcgaattgagaata	2328		
Db	1688	ACATTGAAGTGGACCTGACTGGGAGGCTGCCATCGCTATACAGATCAACTAAAGAATA	1747		
Qy	2329	ttttactcatgttcccaagagtggtgtgtgtgatgaagaacttctacttgtggcttata	2388		
Db	1748	TTTTTGCTCATGTTCCAAAGAGTGGTGTGTGTGATGAAGATCTCTTACTGGTGGCTTATA	1807		
Qy	2389	aagaaatgcacaagctgtgatgagtgagtcaccagtttcatatcatagtagcaagacag	2448		
Db	1808	AAGAATGTCACAAAGCTGTAAATGAGGTGCAGTACAAATTTTCATGTCAGTACCAGACAG	1867		
Qy	2449	tagtacaactgtgtgacataagtgtgaaacaaagtcctcacagagatctctgagagatctg	2508		
Db	1868	TAGTGCATTTGCGGTCATAGTCTGGAAACCAAAATCCTACAAAGTGTCTGAGGACCTTG	1927		
Qy	2509	taagcatatacttgcacctcttaggaccttgcgtgtcttcaatgtacgttttaagcagggc	2568		
Db	1928	TAAGCATACACCTGCCACTCTCTAGAACACTTGTGTGTCTTTCATGTACGTTTAAAGACAG	1987		
Qy	2569	tgggtgctgttccaagactgcataaattgtgtcttttttgaggaccttccaagtagagttac	2628		
Db	1988	TAGGTGCTATTTCAAGACTGCATGAATTTGTGCCCTTTTGCACAGCTTTCAAGTAGAGTCC	2047		
Qy	2629	taqtggaatatccttactcgttgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	2688		
Db	2048	TGTTGGAGTACCCGCTGCGCTGCTTCTTCTGTTGGGTTCAGGTTGTTGCTGAGATGTGGC	2107		
Qy	2689	gaagaaatggactgtctcttattagccagggtgttttattaccagaagatgttaagtgcagag	2748		

Db	2108	
----	------	--

Db	3188	AAAGCTGAGGCGCGCTAGGCTTTTCATCGCCAGAAAGATCATGGCCACAGATGCTCTGCCTTACAGA	3244
QY	3820	aaaaacttcattgaactcataaaactcatgatacaaatacatcagaaaatgcctgggaaag	3879
Db	3248	AAAACCTTCATTGAAACCCACAACTCATGTATGATATACGTACAGAGTAAACAGGGAAGG	3307
QY	3880	aagattccattatgggggaagagagaccccccagtcagtcagtcactactcagaaattgctt	3939
Db	3308	AAGACTTCCATTATGAGGAGAGAGAGACACCTTCAGCAGCTCACTGAGGCCCTCTAGAAATTGCTC	3367
QY	3940	tgggtcctaaacggggtccatctgttacttgaaagaggtgctgacgtgcacatcctttgcc	3999
Db	3368	TGGGCCCTTAAACGGGGCCCGGCTTTTACCGAAAGAGGGTGCTAGACGTGCATCCTCTGCC	3427
QY	4000	agaagaacagagggtgaaatatgaaaaataatgccatggtattatcggcctgtgtccaga	4059
Db	3428	AAGAAGAACAAAGAGGTAAAACTAGAAAATAATGCCATGGTATTGTCAGCATGTGTGCAGA	3487
QY	4060	aatctactgccttaaccccagcacagggggaacccccatagaaactctcagggagaagccctag	4119
Db	3488	AATCCACCGCCCTTAACCCACGACACAGAGGGAAGGCTCTGGACCACCTTAGGGGGAACACTGG	3547
QY	4120	accacttttcattggtatccagactctgcacatagtgaactatatacaggaagctgtgggtcatg	4179
Db	3548	ACCCTCTTTTCATGTGATCCAGACTTGGCACATGGAACTTTATACAGGAAGCTGTGGTCA TG	3607
QY	4180	taatgcacgcagtgctctggcagaagtattttgaagctgtacagctagactctcagcagc	4239
Db	3608	TAATGCATGCAGTGTCTGGCAGAAGTATTTTGAAGCTGTGCAGCTGAGCTCGCAGCAGC	3667
QY	4240	gcattcatgtgacctttttgacttggaaagtggagaatatctttgcctctctttgcaaat	4299
Db	3668	GCATTCACGTAGACCTGTGTGACCTGCAGAGCGCGAGTACCTTATGCCGCTCTGCAAGT	3727
QY	4300	ctctgtgaaactactgtgatcccccatattctctttgaaacctcaaaagataaaacagtgaga	4359
Db	3728	CTCTCTGCAACACTGTCTATCCCATCATCCCTTTGCGCGCCGAGAAGATCAACAGTGA	3787
QY	4360	atgcagatgctctgtcacaactttgacccctggcgcggtggatcacagactgttcttgccca	4419
Db	3788	ATGCGGAGGCTCTGTCTCAACTTTTGACCTTGGCCCGGTGGATACAGACTGTCTTGGCCA	3847
QY	4420	gaatcacggtttataataaagacatgctaagaagagaaaaacc---caattccctatttct	4476
Db	3848	GAATATCGGGTTATATAATAAAGCATGCTAAAGGAGAAAGCCCCACGAGTTCCTGTCTGTG	3907
QY	4477	ttaataaagaattggagattctactttggaggtccattccatccactcctgagttttggcggtg	4536
Db	3908	TTAATCAAGGAATGGGGGATTCAACTTTTGAGTTTTCATTCCTTCCTGAGTTTTGGAGTTC	3967
QY	4537	agctcttcgatataattcaaatagcatcaaggaatggttattctctttggccacaacaa	4596
Db	3968	AGCTTTCGGTGAAATATTCAATAGTATTCAGGAATGGTCATTCCTTCGCCACACAACAA	4027
QY	4597	tttatagaattggattgaaagtgcacctgatgaaaggatctctcagtcagtcctccactgtga	4656
Db	4028	TTTACAGAAATTGGCCTGAAAGTGCCCTCTGTAGAACTAGACCCACGAGTGCCCATGATGA	4087
QY	4657	cctggagacactgcgctttcactatccaggcaattgaaaatctatctgggagatgaaggaa	4716
Db	4088	CCTGGAGCAGCTGTGCGTTTCACCTCCAGGCAATCGAAAACCTTGTGGGAGATGAAGGAA	4147
QY	4717	aacctctgtttgagcactcaaaataggcagcataatggtctgaaagcataaatcagct	4776
Db	4148	AACTCTATTTGGAGGACCTTCAAATATGACAGCATAGCGGTCTGGAAGGCGCTAATGCAGT	4207
QY	4777	ttgcagttgcacagaggttacctgtccctcaggtccctgatcacagaacaactctgtgttcgtc	4836
Db	4208	TTGCGATTGCACAGAGGGCTACCTGCCCTCAGGTCTCTGATACACAAACATCTGGCTCGCG	4267
QY	4837	ttctatagttgttcttcctaacaataaatacagaagatacacacctgctcttctgtctatag	4896
Db	4268	TCCTGTCACTTATCTCTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTT	4327

Qy	4897	atctgtttcaatgttttttgggtgctgtgttagcattccactctctgttatctgggatgacc	4956
Db	4328	ATCTCTTCATCTTGTGTGCGCGCAGCTTTAGCGTCCATCTCTTATTTGGATGACA	4387
Qy	4957	ctgttgatctgcagcctctctcaagttagttctctctaataacacacttatctcttccatt	5016
Db	4388	CGGTGGATCTGCAGCGCTGCGCCACTTAGTCTTCATATAACCACTCTATCTCTTCCATC	4447
Qy	5017	tgatcaccatggcacacatgcttcagatctcagatctactactlacagtacaca-----cag	5064
Db	4448	TCATCACCATGGCGCACATGCTTCAGATACCTCTTACAACACATACAGATCTGTCTCCAG	4507
Qy	5065	gocctaccctctgctcaggttccaagaagacagtgaagagctcattccgcgactctttctt	5124
Db	4508	GGCCGCCGCTGCTGAGGTGAAGAGGATAGTGAGGAGGCTCGCTGTGCATCTGCTTTCT	4567
Qy	5125	ttgcagaataattctcaatatacaaagtgcctcoatgggttgatattcctgcgtcggtatt	5184
Db	4568	TTGTGGAAAGTGTGCGACGACACACAGACGGCCTCAC'TGGGTGCGGTGCTCCCGCTG	4627
Qy	5185	tgtgggtctcactgaagaatggcatcaccccttatctctcgtgctgctcatgtttctcc	5244
Db	4628	TGTGGCTCTCCTTGAGGAACGSCATCACCCCTTACTTCGCTGTGCTGCAC'TCTTTC	4687
Qy	5245	actattacttgggtgaactccgcctcaggaactgcatacccaattctgcagaaggagat	5304
Db	4688	ACTATTATTACTTGGAGTAGTCCGCGCTGAAGAACTGTTTGCCAA'TTCTGCTGAAGGAGAAT	4747
Qy	5305	acagtgcactctgtagctatctattctacctacaaaattgtctcgtcttccaggaa	5364
Db	4748	TCAGTGCAC'TCTGTAGCTATCTATCTTTACCCACAAATTTGTCCTGCTTTCCAGGAAT	4807
Qy	5365	attgggtactgtaaagcccttgcctccagagtggtgtgcagatcctgccttactaaact	5424
Db	4808	ATTGGGATACCATAAGGCCCTTACTACAGAGCTGGTGTGGAGATCCTTGCTTACTCAAGT	4867
Qy	5425	gtttgaagcaaaaaaacccgtgggtcaggtacccctagaaaaaagaaatagtttgatagac	5484
Db	4868	CTTTGAAGCAGAAAAAGTGTGTGGTACAGTACCTAGAAAAAGAAATAGTTTGATAGAC	4927
Qy	5485	ttcctgtagactatagctgcctccctgaatcaagcttctcatcagggtgccacgctcgt	5544
Db	4928	TTCTTGAGGACTACAGCTGTCTCTAAATCAGGCTTCTCAC'TTTAGGTGTCCACGGTCTG	4987
Qy	5545	cagatgtagcgaaagcatcctgctcctgccttcttctgtggggctatactactatgtctc	5604
Db	4988	CAGATGATGAGCGAAAGCATCTGCTCTGTCTTTCTGTGTGGGGCCATCCTGTGTCTTC	5047
Qy	5605	agaacattgtgctccaggaataattgtgaacgggggaagaggttgagagctgtcattttcacg	5664
Db	5048	AGAAATCTGTTGCCAAGAAATAGTGAATGGGGAAGAGGTTGGAGCGTGGCTTTTTCATG	5107
Qy	5665	cacttcactgtgaacccgagctcacatttctcctaaaaatacagaagaatccgagtgctcc	5724
Db	5108	CGCTTCATTGTGGTGTGGAGTCTGCATTTTCTTAAATCCGAGAA'TCGAGGGTGGTCC	5167
Qy	5725	tgtttgaaggtaaagccagagcgtgtgcctatccagctccttacttggtatgaataaggag	5784
Db	5168	TGGTGGAAAGGAAAGCCAGAGGCTGTGCCTTACCCAGCCCCCTTACTTGGATGAATATGGAG	5227
Qy	5785	aaacagaccctggcctgaagaggggcaaccccttcaattatctctgtgagcggtatcgga	5844
Db	5228	NAACAGACCCAGGGCTAAAGAGAGAAACCCCACTTCATTTATCTCGGGAGCGGTATCGGA	5287
Qy	5845	agctccatttggctctggcaacaacactgcattatagaagagattgctagagaccagaga	5904
Db	5288	AGCTGCATTTGTCTGCGCAACAGCACTGCTATTATAGAGAGATTTGCTCGGAGCCAGAGA	5347
Qy	5905	ctaactcagatgttatttggattcactcagctggcagttactgtgactcacaactctgcctcaag	5964
Db	5348	CTAATTCAGATGCTATTGTGATTTAACTTGCAGTTACTCTGAGCTTCAGTTCCTCCCTCAAG	5407

Qy	5965	acaa	tcacaaa	tgacgacag	tagtaag	gcgctgattcaaaa	tattgga	aaactttctgag	6024
Db	5408	ACAATCATGAGTGACATCAATAATAAAGACTGATCTAAAAATTC	TAGAGA	ACTTTCTGAGG	5467				
Qy	6025	gctgggaag	tattgagag	gctcttttgc	tccatgtccaggttc	cacttacatca	ataa	aaat	6084
Db	5468	ACGGGGGAAGTATTGGAGGGTC	TTTTGATCCATGTCCAGATT	CACACACATTAATAAAAT	5527				
Qy	6085	atttctaat	gtagtattgcttttca	atagcaacacat	atgcttcacaggg	aaaaa	-aggac		6143
Db	5528	ATTCTTAAATGGNAATATTGCTTTCA	ATTCAACATAGCTTCA	AGGGGAAACAGAC	5587				
Qy	6144	atagatca	aatctgt	-----	-----	-----	-----	-----	6160
Db	5588	ATAGATTAAATGTTTATGTTCTAGACAC	ATAAGAAATGCTGTT	CATCCAAAGTGCTCAT	5647				
Qy	6161	atgtgctag	tattccagga	atttattcc	ccctc	ataattgtctc	taatttc	taatttt	6220
Db	5648	TTTCTGCTAATATTTCAGAAAACTCCTTTCCCTT	TCATAAC	TGCTAGTTCATTTTCAT	5707				
Qy	6221	catccact	tgtagatgaag	tcacgc	caaacag	tgttagac	atttattg	tgttggttaac	6280
Db	5708	CACCCACCTGGTTAATGAGGTACATTAAC	ATTG	TGGACATTTCTCCATCTG	6280				
Qy	6281	tcttctgca	attttgtattg	tgtg	tgtt	6307			
Db	5768	ATCTCTGCACCTTTGATTGTTGGT	TTT	5794					
RESULT	2								
LOCUS	AR121463	6395 bp	DNA	PAT	16-MAY-2001				
DEFINITION	Sequence 1 from patent US 6159732.								
ACCESSION	AR121463								
VERSION	AR121463.1	GI:14105039							
KEYWORDS	Unknown.								
SOURCE	Unknown.								
ORGANISM	Unclassified.								
REFERENCE	1 (bases 1 to 6395)								
AUTHORS	Varshavsky,A. and Kwon,Y.Tae.								
TITLE	Nucleic acid encoding mammalian Ubr1								
JOURNAL	Patent: US 6159732-A 1 12-DEC-2000;								
FEATURES	Location/Qualifiers								
source	1..6395								
BASE COUNT	1802 a 1376 c 1519 g 1698 t								
ORIGIN									
Query Match	70.3%;	Score	4435.4;	DB 6;	Length	6395;			
Best Local Similarity	86.3%;	Pred. No. 0;							
Matches 4997;	Conservative	0;	Mismatches	721;	Indels	69;	Gaps		
Qy	590	ttcaggggc	gcgtcgtaaaag	tgcgtccctgtctc	-tccgaccg	ggccacaggttccgct	648		
Db	8	TTCAGGGGGCGTCGTAAAAGTGTGTC	CCCTGTCCGTCGGCGCCG	CACAGGTTTCGCCT	67				
Qy	649	tgctctg	gcgggggtcg	gcaactgcag	cgctcagttccctc	caagatggc	gacgag	708	
Db	68	AGCTGCGCGCGGGGTGCGGAAC	TGCGGGCGTTCGTTTCCCTTTA	AGATGSCGACG	GAAG	127			
Qy	709	aggctg	gaggtactag	agagatg	aaatac	gcgcgagg	tacc	ccccagccctc	768
Db	128	AGATGACGCGCGGAGAGGATG	AGCTGACGCGGAGCCTTCCCTT	CCCTTTAAGATGSCGACG	CGGC	187			
Qy	769	tggcatct	tggtggtatc	cacga	gttgatttttata	ctgcttctt	cgatcattt	ggcac	828
Db	188	CGGCATCTG	TGGGTGGAATACAA	AGTGAATTCAT	ACTGCTTTCTTACAT	CTTTGGCAC	247		
Qy	829	aattg	tgcagaaa	tatttact	tgtgaa	atggacc	cagactt	ggaaaag	888
Db	248	AATTACTG	CCAGAAATTTATTTT	GCTGAGATG	ACCCAGATTTT	CGAAAAGCA	AGAGAGA	307	

YNHLPLHLITMAHMLIILLTTDTDLSPGPPLEGEEDSEBARCASAFFVEVVSQHTDG
LTCCGAPCWILSLRNGITPYLRCAALLFHYLLGVAPPPELFPNSAGRESALCSYL
SLPTNLFLEQEXYWDITIRPLLRQWGDPAKLSKQSAVVYRPRKRNLSLELEPEDYS
CLINOASHFRCPSADDERKHPVLCLFCGAILCSQNTCCQEIIVNGEEVGCACVFHALHC
GAGVICFLKRECRVILVEKGARGCAYPAPYLDYGETDPLGKRGNPLHLRSRERYKL
HLVMOQHCIITEIARSQETWQMLFQFNQOLL"
5389..6395
/gene="ubrl"
5429..5434
/gene="ubrl"

polyA_site

BASE COUNT 1802 a 1376 c 1519 g 1698 t

ORIGIN

Query Match 70.3%; Score 4435.4; DB 10; Length 6395;
Best Local Similarity 86.3%; Pred. No. 0;
Matches 4997; Conservative 0; Mismatches 721; Indels 69; Gaps 6;

Qy 590 ttcaggggcccgtctaaagtctgcctcgtctc- tccgaccggccacaggtttccgct 648
Dy 8 TTCAGGGGCCGCTGTAAGGTGCTCCCTGCTCGCGTCGGGCGGCCACAGGTTTCCGCT 67
Qy 649 tgcctctggcggcggtcgcaactcagcgctcagtttccctcaagatgcggacgag 708
Dy 68 AGCTGGCGCGCGGGGTGGGAACCTGCGGGGCTTCGTTTCCCTTAAGATGGCGGACGAAG 127
Qy 709 aggcctggaggctactgagaggatggaatcagcgqggagttaccgccagccctcagctc 768
Dy 128 AGATGGAGCGGCCGAGAGGATGGAGCTCAGCCGGAGCCCTCCCTGGCCCGCAGCGGC 187
Qy 769 tggcacttggggatcagcaagtgtgatttttatactgcttcttgcatactttggcac 828
Dy 188 CGGCATCTGTTGGGATCAGCAAGTTGATTCTTACTACTGCTTCTTACATCATTTGGCAC 247
Qy 829 aattgggtccagaatttacttctgaaatggaccceagacttggaaagcaggagaaa 888
Dy 248 AATTAGTCCCAAGATTTATTTTGTGTAGATGGACCCAGATTTGGAAAGCAAGAGAGA 307
Qy 889 gtgtacaaatgtcaatttactccactcagtggaatgtacttatttggagaagatccagata 948
Dy 308 GTGTACAGATGTCAATCTACTCTCTTTGGAGTGGTACTTATTTGGAGAGGATCCGGATA 367
Qy 949 ttgtctagagaaattgaagcacagtgagcatcttcagctttgtggggggttttcaaaa 1008
Dy 368 TTGTCTTAGAGAAATTAACACACAGTCAGCGCTTCCAGTGTGTGGGAAGGTTTTCAAAA 427
Qy 1009 gtggagagacaacctattcttgcagggtattgtgaattgatcccaacatgtgtactctgta 1068
Dy 428 GTGGAGAAACAACATATTTCCTGTAGGGATTTGTGCAATTTGATCCCAACGTGTGTCTGTA 487
Qy 1069 tggactgctccaggacagtggttcataaaaaatcactgttacaagatgcatacttctactg 1128
Dy 488 TGGACTGTTTCCAAAGTAGTGTTCATAAAACCATCGTTACAAAGATGCATCTTACTGT 547
Qy 1129 gagagggttctgtactgttgagacacagaggcattggaactggccctttttgtgtaa 1188
Dy 548 GAGGGGCTTCTGTACTGTGGAGACAGAACGCTGGAAAACTGGCCCTTTTGTGTGG 607
Qy 1189 atcgtaaccttggagagacaggtactataaaagagaattcaccgtgtccgttgtaatgaag 1248
Dy 608 ATCAGGACCTTGGAAAGACAGGTACTACAAAAGAGAGGCTTACATTTGCCCATTTGAATGAAG 667
Qy 1249 aggttaattgtccaagccaggaaaaatttcttctcagtgataaaaatgtcgtagaaaatga 1308
Dy 668 AGGTGATTGTCTCAACCCAGGAGAAATTTCCCTTCCGTTGATGATAAAATACATTTGAGAAATGA 727
Qy 1309 ctatatgtggaagagaaaaaactgcctcctgaaactccagataaaggagagaaaaatgaaa 1368
Dy 728 CTATATGGGAAGAAAGAAATTTGCCTCTCTGTAACCTCAGATAAGGGAGAAATGAAAC 787
Qy 1369 gatactattgtctcttccatgatgaaacacacattcattgaccacgtcatatacagcc 1428
Dy 788 GATACTATTGTGCTTTTCAACGATGAGACCAATTCGTATTGATCATGTGATCTACAGTC 847

Qy 1429 tacaagaagactcttgaactgtgagctcgagagccaggttgatataccactgcccattgaca 1488
Dy 848 TGCAGAGAGCTCTAGATTGCGAGCTTTCAGAGCAGCAGGTGTCACAGACTGCGCATCGACA 907
Qy 1489 aagagggctcgtcggcgtgttaaagcggagcttactctgtccaggaagcaaaagaag 1548
Dy 908 AAGAGGGTCGCCGGGCTGCAAAAGCAGGTGTGTATGCCACTTCCAGGAAGCAAGGAGG 967
Qy 1549 atataaagagtcattcagaaaaatgtctctcaacatccacttcattgtagaagtattacact 1608
Dy 968 ATATAAAGAGTCACTCAGAGAACGCTCTCTCAGCACCCCCCTCCATGTGGAAGTGTCTGCAC 1027
Qy 1609 cagagattatggctcactcagaaaaatttgcttgcgtcttggcttctcctggatgaacaaatta 1668
Dy 1028 CCGTGGTTATGGCTACCAGAAATTCGCTCTGCGCCTTGGCTCTGGATGAACAAATAA 1087
Qy 1669 tgagctattcaagtgaacttttaggcagatcttttgcgaagcatgccttagagaagaacctg 1728
Dy 1088 TGAGCTATTCAAGTGACTTTAGACAGATATTTTGCAGGGCTGCCTCGTGAAGAGAACCTG 1147
Qy 1729 actcggagaatccctgtctcataagcaggtttaatgttttggatgcgaagctttataaag 1788
Dy 1148 GCTCTGAAAATCCCTGCTTATTAAGCAGACTAATGCTTTGGGATGCAAACTTTATAAAG 1207
Qy 1789 gtgcccgttaagatcccttcataatgatcttcagcaggtttttttatggagatggaataca 1848
Dy 1208 GTGCCGTGAAGATCCCTTCATGAATTCATCTTTAGTAGTTTTTTTATGGAGATGAATACA 1267
Qy 1849 aaaaactcttgcattggaatttggtaagtattataaaaaactgcagaagaataataca 1908
Dy 1268 AAAAATCTTTTGTATGGAATTTGTGAAGTATTAATAACAACGCAAGAAAGAGTACATCA 1327
Qy 1909 gtgatgatcatgacagaagtatctataactgcacttccagttcagatgttactgttc 1968
Dy 1328 GGCAGCACACGAGAGAGCATCTCCATACCCTCTGTCGTCGAGATGCTCACCGTCC 1387
Qy 1969 ctactctgctgcagactcttattgaagagcagaatgttactctctgtcattactgaaact 2028
Dy 1388 CGACCTTGGCCCGGCATCTTATTGAAGAGCAGAAATGTTATTCTGTCTATTACTGAAACGC 1447
Qy 2029 tctagaagattttacactgagcttgcagaggaacaaataaactcaactccaggggtata 2088
Dy 1448 TGCTAGAAGTTTTACCTGATACTTGGACAGAACATAAATAATTCACCTCCAGGGTTATA 1507
Qy 2089 gccagacaaattggagaagtataatgcagtaatatgtgacctaaagtatatctcgtatca 2148
Dy 1508 GCCAGGACAACTGGGAAGAGCTACGAGTTTATGTGACCTTAAGATATATCTTGATTA 1567
Qy 2149 gcaaaccccaaatatggacagaagaattgaagaatgcagttccttgaagggttttcgactct 2208
Dy 1568 GCAAGCTCTGTCAATATGGACAGAACGATTAAGAGCGCAGTTTCCCTGGAAAGTTTCCGGTCTT 1627
Qy 2209 ttttgaagattcttactctgatagcaggggaatgaagaataccgagaagacaggttgggcaac 2268
Dy 1628 TTCTGAAGATTCTTACCTGTATGTCAGGGAATGAAGAAATCAGAAGAACAAAGTTGGACAAC 1687
Qy 2269 acattgaagtggatcctgattgggaggtccttgcatacacagatgcaattgaaagata 2328
Dy 1688 ACATTGAAGTGGACCTGACTGGGAGGCTGCCATCGCTATACAGATGCACACTTAAGAAATA 1747
Qy 2329 ttttactcatgtttcccaagagtggtgtgtgtgatgaagaactcttacttgtggtctata 2388
Dy 1748 TTTTGTCTCATGTTTCCCAAGAGTGTGTGTGTGATGAAGATCTCTTACTGTGTGCTTATA 1807
Qy 2389 aagaaatgcacaaagctgtgtagtgaggtaccagtttccatctcagtagcaagaagcag 2448
Dy 1808 AAGAATGTCACAAAGCTGTAATGAGGTGCGAGTACAAAATTTTCATGTCCAGTACCAAGACAG 1867
Qy 2449 tagtacaactcgtgtgacatagtttggaaaacaaagctccacagatctcagagatcttg 2508
Dy 1868 TAGTGAATTTGTGCGGTTCATAGTCTGGAAACCAAAATCCTACAAAGTGTCTGAGGACCTTG 1927

Qy	2509	taagcatacatctgccaactctctcaggacccttgctgggtcttcatgtacgtttaagcaggc	2568
Db	1928	TAAGCATATACACCTGCGCACTCTCTAGAACAACACTTGTCTGGTCTTCATGTACGTTTAAGCAGAC	1987
Qy	2569	tgggtgcgttttcaagactcgatgaatttgtctcttttttaggaccttcaagttagagttac	2628
Db	1988	TAGTGTCATTTCACAGACTGCATGAATTTGTGCCCTTTTGACAGCTTTCAAGTAGAGGTCC	2047
Qy	2629	tagtggaaatactcttttacggttgtctgggtgtgtgtgtgtgccagggtgtgtctgagatgtggc	2688
Db	2048	TGGTGGAGTACC CGCTGCGCTGCCCTGGTCTCTGGTGSCTCAGGTGTTGTCTGAGATGTGGC	2107
Qy	2689	gaagaataaggactgtctotattatagccagggtgttttataccacaagatgttaaagtcagag	2748
Db	2108	GAAAGAACGGGCTCFACATCAFTCACCGCAGGTTTCTATTATCAAGATGTATAATGACAGG	2167
Qy	2749	aagaaatgatgataaagatatcatcatgcttcagattgtgcattgtgcattcttaatggatcccc	2808
Db	2168	AGGAANTGTACGATAAAAGATATCATCATVGTCTCAGATTGGAGCATCTATTAATGGATCCCCA	2227
Qy	2809	ataagttcttctactgttacttcagaggtatgaacttgcggaggtgccaggctttttaacaagacca	2868
Db	2228	ACAA GTTCTTGTACTGTCTTACAGAGATATGA ACTTACTGATCTGCTTTTAA CAAGACCA	2287
Qy	2869	tatctacaaaagaccagatttgtattaacaacata taatcacataatagagaanaatgcttc	2928
Db	2288	TATCCAAAAAGACCAGGATTTTGATTAACACGTATATAACATTAATAGAGAANAATGCTTTC	2347
Qy	2929	aggtccctcatctatgttgggtgagcgttatgtaccctggagtgagtggaatgtgaccaaag	2988
Db	2348	AGGTCTCATCTATATTGTGGGAGAACGTTATGTACTCGGTGGGAATGTTTACCAGAG	2407
Qy	2989	aagaggtcacaatgagagaaatcattcacttgccttgcatttgaa cccatgccacacagtg	3048
Db	2408	AGGAGGTTAATGAGAGAGATTACTACTTACTTGTGATTGAGCCCATGCCACACAGTG	2467
Qy	3049	ccattgcccataatttacctgagaatgaaaa taa tgaactggcttagagaatgtcataa	3108
Db	2468	CCATGCGCCAGAAA CCTACTCGAGACGAGAAAATATGAAACTGGCTTAGAGAATGTCATAA	2527
Qy	3109	acaa gtggccacattttaagaa accaggtgtatcaggccatggagtttatgaactaaaaag	3168
Db	2528	ACAA GTGGCCACATTTAAGAAACCAGGTTGTGCGGCCATGGAGTTTATGAATTTAAAG	2587
Qy	3169	atgaatcactgaagacttcaatatgtacttttattcattactccaaa acccagcatagca	3228
Db	2588	ATGAATCACTGAAAGACTTCAATATGTACTTTTACCATTATTCTTAA AACACACAGCATAGCA	2647
Qy	3229	aggctgaacat atgcagaagaaaaggaga aaacaaga aacaaaagatgaagcattgccgc	3288
Db	2648	AGGCTGAACATATGCAGAAGAAAAGAGAAAACAAGAAAATAAAGATGAAGCATTGCCGC	2707
Qy	3289	caccacacctccctgaattctgcctgctttcagcaaa gtagtaaaccttctcaactgtg	3348
Db	2708	GGCCACCTTCCTCCAGAGTTCTGCCCTGCTTTTCAGCAAAGTAGTCAACCTGCTCAGCTGTG	2767
Qy	3349	atatcatgatgtacattctcaggaccgttattggcgggcaatagacacagattcttaact	3408
Db	2768	ATGTTATGATATACATCTTCAGGACCATCTTTGAGGGGCGCATGGACACGGAGCTTAATC	2827
Qy	3409	tgtggaccgaagggaatgtccaaatggcttttcatattctggcattgggttttactagaag	3468
Db	2828	TGTGGACAGAAGGGATGCTGCAGATGGCGTTC CATATATTTGACACTGGGCTTGCTGGAAG	2887
Qy	3469	agaagcaacagcttcaaaaagctcctgaagaaga agtaaacatttgacttttatcaataagg	3528
Db	2888	AGAAGCAGCAGCTTTCAGAAAGCTCCTGAAGAGGAAGTGGCTTTTGTACTTTTACCATAAAG	2947
Qy	3529	cttcaagattgggaagttcagccat- ----- -gaatatacaaatgcttttggaaaaaac	3579
Db	2948	CTTCAAGATTGGGAAGTTTACGCCATTGAATGTCACGA AATATACAAATGCTCTTGGAAAGAC	3007
Qy	3580	tcaaaggaaattcccaggttagaagcccaagaagacatgataacgtgga lacticagatgt	3639

[illegible]

Db 4088 CCTGGAGCACGTGTGCGTTCCACCATCCAGGCAATCGAAACCTGTTGGGAGATGAAGAA 4147
Qy 4717 aacctctgttbbgagcaacttcaaaataggcagcataatgtgtcgtgaaagcattaatgcagt 4776
Db 4148 AACCTCTATTTGGAGCACTTTCAAAATAGACAGCATAGCGGTCTGAAGGCGCTAATGCAGT 4207
Qy 4777 ttgcagttgcaagagattacctgtcctcaggtccctcgtatcacagaacaacatcgtgttcgc 4836
Db 4208 TTGCAGTTGACAGAGGGCTACCTGCGCCTCAGGTCCTGATACACAAACATCTGGCTCGGC 4267
Qy 4837 ttcctcagttgttctcctcaataaataacagaagatcacaccatgcctctcgtctctatag 4896
Db 4268 TCCTGTCACTTATTTCTTCCCTAACCTGCATACAGAAATACACAGGCGCTTCTGTCTGTGG 4327
Qy 4897 atctgtttcatgttttgggtgggtcgtgtgttagcatctcccatcctctgtattgggagatgacc 4956
Db 4328 ATCTCTTCCATGTTCTGTGTGGCGCAGCTCTTAGCGTTCCCATCTCTTATTGGGATGACA 4387
Qy 4957 ctgttgatctgcagcctttcttcagttagttcttctcctataaacaacatttatctctccatt 5016
Db 4388 CCGTGGATCTGACGCCGTGCGCACTTAGTTCTTCATATAACCACTCTATCTCTTCCATC 4447
Qy 5017 tgcacacatggcacacatgcttcagatactactactacagtagaca-----cag 5064
Db 4448 TGATCACATGGCCACATGCTTCAGATACTCTTACACAGATACAGATCTGTCTCCAG 4507
Qy 5065 gcttacccttgcaggttcaagaagacagtgaagagctcattccgcattctcttctct 5124
Db 4508 GGCCGCCGCTTGCTGAGGCTGAAGAGGATAGTGAGGAGGCTCGTGTCATCTGCTTTCT 4567
Qy 5125 ttgcagaattctcaatatatacagaagtggtccatctgggtgtgatattcctgtggtgatt 5184
Db 4568 TTGTGGAAGTGCAGCAGCACAGACGCGCTCACTGGGTGGCGGTCTCCCGCTGGTACC 4627
Qy 5185 tgtgggtctcactaagaatggcatccctctattctcgtgctgctcattgttttcc 5244
Db 4628 TGTGGCTCTCCCTGAGGAACGGCATCACCCCTTACCTCGGTGCTGCATCTGCTTTCC 4687
Qy 5245 actatttacttgggttaactccgctgaggaactgcataaccaatctgcagaagagagt 5304
Db 4688 ACTATTTACTTGGAGTAGCTCCGCTGAAGAAGTGTGGCCAAATCTGCTGAAGGAGAT 4747
Qy 5305 acagtgcactctgactatctatttaccatacaaaatttgttctcgtctcttccaggaat 5364
Db 4748 TCAGTGCACCTGTAGCTATCTATCTTTACCCACAAATTTGCTGCTTTTCCAGGAAT 4807
Qy 5365 attggatactgtaagcccttgcagaggtggtgtgcagatcctcctactactaaact 5424
Db 4808 ATTGGGATACCATAAGGCCCTTACTACAGAGGTGGGTGGAGATCCTGCCCTTACTCAAAT 4867
Qy 5425 gtttgaagcaaaaaaacaccgtgtcaggtaccctcctagaaaaaataatgattgtatagac 5484
Db 4868 CTTTGAAGCAGAAAGTCTGTGTCAGTACCTACCTTAGAAAAAGAAATAGTTGATAGAGC 4927
Qy 5485 ttctgtgactatagctgtcctcctgaatcaagcttctcatttccattcaggtgccacggtctg 5544
Db 4928 TTCTCGAGGACTACAGCTGTCTCTTAATCAGGCTTCTCACTTTAGGTGCCACGGTCTG 4987
Qy 5545 cagatgtagcgaagacatcgtcctcctgtccttctgtggtggtcctatactatgtctc 5604
Db 4988 CAGATGATGAGCAAAAGCATCTGTCTCTGTCTTCTTGTGGGCCCATCTGCTGTCTC 5047
Qy 5605 agaacattgtgccagaaattgaaacggggaagaggttggaactgtgcatttttccag 5664
Db 5048 AGAACATCTGTGCCAAGAAATAGTGAATGGGAAAGAGGTGGAGCGTGCCTTTTTCATG 5107
Qy 5665 cacttcactgtgagccggagctcgtcatttctccttaaaatacagagaatgcgcagtggtcc 5724
Db 5108 GCCTTCATTGTGGTCTGGAGTCTGCATTTCTTCTAAAATCCGAGAAATGCAGGGTGGTCC 5167
Qy 5725 tgggtgaaggtaaagccagaggtgtgctatccagctccttacttgatgaatgag 5784
Db 5168 TGGTGAAGGAAAAAGCCAGAGCTGTGCTTACCCAGCCCTTACTTGGATGAATATGGAG 5227

Qy 5785 aaacagaccctggcgtgaagaggggcaaccccttcaatttatctctgagcggtatcgga 5844
Db 5228 AAACAGACCCAGGGCTAAAGAGAGAAACCCACTTCTATTATCTTCGGGACGGTATTCGGA 5287
Qy 5845 agctccatttggctcggcaacaacactgctattatagaagattgctagagccaaagaga 5904
Db 5288 AGCTGCATTTGGTCTGGCAACAGCACTGCTATTATAGAAGATTGCTCGAGGCCAGAGA 5347
Qy 5905 ctaatcagatgttatttgattcaactggcagttactgtgagctccaactcctgcctcaag 5964
Db 5348 CTAATCAGATGCTATTGGAATTAAGTGGCAGTTACTCTGAGCTTCAGTTCTGCTCAAG 5407
Qy 5965 acaatcaaaaaagcagcagtagtaaaagcgtgattcaaaatattgaaaaacttctcgagg 6024
Db 5408 ACAATCATGAGTGATCAATCAATAAAGACTGATCTAAAATTTCTAGAGAACTTTCTGAGG 5467
Qy 6025 gctggaaaagattgaggggtcttttgcctgctccaggtccaggttcaactacatacaataat 6084
Db 5468 ACGGGGAAGTATTGAGGGTCTTTTGATTCATGTCAGATTCCACACATTAATAAAAT 5527
Qy 6085 attcttaatggagctattccttcaattagcaaacatactgttcacaggaataaa-aggac 6143
Db 5528 ATTCTTAAATGAATATTGCTTTCAATTATCAACATAGCTTCAGGGGAAACAAAGAC 5587
Qy 6144 atagatcaatctgt-----ttt 6160
Db 5588 ATGATTAATGTTTATGTTCTAGAACACATAAAGAAATGCTTGTTCATCCAAAGTGCTAT 5647
Qy 6161 atgtcctagatttccaggaatttattcccttcaataattgtctcatttctatttattt 6220
Db 5648 TTCTCTAATATTCCAGAAAACTCCTTTCCCTTCATAACTGTCCCTAGTTCAATTCATAT 5707
Qy 6221 catcacttgtagatgaagtcacgtcaaacaggtttagacatttattgtgttggttaac 6280
Db 5708 CACCACCTGGTTAATGAGGTCACATTAGCATTTGTGCACATTTCTCCATCTGGCTAAC 5767
Qy 6281 tcttctgcaatttgtattgttggttt 6307
Db 5768 ATCTCTGCACCTTGTATTGTTGGTGT 5794

RESULT 4

AK027803
LOCUS Homo sapiens CDNA FLJ14897 fis, clone PLACE1004743, weakly similar to PROBABLE N-END-RECOGNIZING PROTEIN.
DEFINITION AK027803
ACCESSION AK027803
VERSION AK027803.1 GI:14042751
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens placenta cDNA to mRNA, clone_lib:PLACE1
clone:PLACE1004743.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Aotsuka,S., Yoshikawa,Y., Matsunawa,H., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuho,Y. and Sasaki,N.
NEDO human cDNA sequencing project
TITLE Unpublished
JOURNAL 2 (bases 1 to 2550)
REFERENCE Isogai,T. and Otsuki,T.
AUTHORS Direct Submission
TITLE Submitted (10-MAY-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 252-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)
JOURNAL NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection: Helix Research Institute (supported by Japan Key Technology Center etc.)
COMMENT

[illegible]

Query Match	40.4%	Score 2546.8	DB 9	Length 2550
Best Local Similarity	99.9%	Pred. No. 0		
Matches 2548	Conservative 0	Mismatches 2	Indels 0	Gaps 0
Qy 1189	atcatgaacctggaagcagcagggtactataaagaagaattcaacgctgcgtccgttgaaatgaag	1248		
Db 1	ATCATGAACCTGGAGAGCAGGTACTATAAAGAGAATTACGCGTGTCCGTTGAATGAAG	60		
Qy 1249	aggrtaattgtcccaagccaggaataattcttcctcagtgataaaatatgtcgtgagaatatga	1308		
Db 61	AGGTAATTGTCCACAGCCAGGAAAAATATTTCCTTCAGTGTAAAAATATGCCGTGAAATGA	120		
Qy 1309	ctatatgggaaggaagaaagaaactgcctcctgaactccagataagggagagaaaaatgaaa	1368		
Db 121	CTATATGGGAAGAGGAAAAAGAACTGCTCCTGAAC7CCAGTAATGAAGGAGAAAAATGAAA	180		
Qy 1369	gatactattgtccttttcaatgatgaacaccattcatatgaccacgtcatatcacagcc	1428		
Db 181	GATACTATTGTCTTTTCAATGATGAACACCACTCATATGACCACTCATATACAGCC	240		
Qy 1429	tacaaagactcttgactgtgagctgcagaggcccgattgcataccactgccaattgaca	1488		
Db 241	TACAAAGAGCTCTTGACTGTGAGCTCGCAGAGGCCCAGTTGCATACCACCTGCCATTGACA	300		
Qy 1489	aagagggtcgtcgggcgtgttaaagcgggagcttatgctgcttgcaggagaagcaagaag	1548		
Db 301	AAGAGGTCGTCGGCGTCTTAAACGGGAGCTTATGCTGCTTGCAGGAAGCAAGGAAG	360		
Qy 1549	ataaaagagtcatcagaataatgctctcaacatccacttcagtagaagtattacact	1608		
Db 361	ATATAAAGAGTCATTACAAAAATGCTCTCAACATCCACTTCATGTAGAGTATTACACT	420		
Qy 1609	cagagattatggctcatcagaaaaattgcttttgogtcttggttcctcgtgatgaacaaaaatta	1668		
Db 421	CAGAGATTATGGCTCATCAGAAAAATTGCTTTGCGTCTTGGTTCCTGGATGAACAAAAATTA	480		
Qy 1669	tgaactattcaagtgaacttaggcagatcttttgcgaagcatgaccttagagaagaacctg	1728		
Db 481	TGAGCTATTCAAGTGACTTTTAGCCAGATCTTTTGCCAAAGATGCTTTAGAGAAGCAACTG	540		

[illegible]

Qy	2809	ataagttcttctgacttcagaggtatgaacttgcgaggttttaacaagacca	2868
Db	1621	ATAAGTCTTGTACTGTGACTTCAGAGGTATGAACCTTGCCGAGGCTTTTAACAGACCA	1680
Qy	2869	tatctacaaaagaccagatttgattaaacaataataacactaataagaagaatgcttc	2928
Db	1681	TATCTACAAAGACCAGGATTGTATTAACAATAATAACACTATATAGAGAAATGCTTC	1740
Qy	2929	aggtctctcatctatatgttggtgagcgttatgtacctggagtggaagtgtgaccaag	2988
Db	1741	AGGTCTCATCTATATTGTGGGTGAGCGTTATGTACCTGGAGTGGGAAATGTACCAAG	1800
Qy	2989	aagagtcacaaatgagagaaatcattcactgtcttgattgaaccccatgccaacagtg	3048
Db	1801	AAGAGTCACATGAGAGAAATCATTTACTTGTCTTGTGATTGAACCCATGCCACACAGTG	1860
Qy	3049	ccattgcaaaaatttacctgagaatgaaataatgaactggcttagagaatgtcataa	3108
Db	1861	CCATTGCCAATAATTTACTTGAGATGAATAATTAAGAACTGGCTTAGAGAAATGCTATAA	1920
Qy	3109	acaaagtggccacatttaagaacaccaggtgtatcgagggccatggagtttatgaactaaaag	3168
Db	1921	ACAAAGTGGCCACATTTAAGAAACACAGGTGTATCAGGCCATGGAGTTTATGAACATAAAG	1980
Qy	3169	atgaactcactgaagaactcaatgtacttttaccatttaccataccaaacccagcatagca	3228
Db	1981	ATGAATCTACTGAAGACTTCAATATGTACTTTTATCATTTACFCCAAAACCCAGCATAGCA	2040
Qy	3229	aggtcgaaacatgcagagaaaagagaaacaaagaaagaaagaaagaaagaaagaaagaaag	3288
Db	2041	AGGCTGAACATATGCGAAGAAAGAGAGAAACAGAAACAGAAACAGAAATGAAGCATTTGCCGC	2100
Qy	3289	caccaccacctctgaattctgcctgtcttcagcaaaagtattaaccttctcaactgtg	3348
Db	2101	CACCACCACCTCTGAATTTCTGCCCTGCTTTCAGCAAAAGTGATTAACTTCTCAACTGTG	2160
Qy	3349	atatcatgatgtacattctcaggaacgttatgttagcgggcaatagacacagattctaaact	3408
Db	2161	ATATCATGATGTACATTTCTCAGGACCGTATTTTGAGGGGCAATAGACACAGATTCTAACT	2220
Qy	3409	tgtgaccgaaggatgtccaaatggcttttcattctgcatgtgggttttactagaag	3468
Db	2221	TGTGACCGAAGGATGCTCCAAATGGCTTTTCATATTTCTGGCATTTGGGTTTACTAGAAG	2280
Qy	3469	agaagcaacagcttcaaaaagctcctgaagaagaagtaacatttgacttttactataagg	3528
Db	2281	AGAAGCAACAGCTTCAAAAAGCTCCTGAAGAAGAGTAACATTTGACTTTTATCATAGG	2340
Qy	3529	cttcaagattgggaagttcagccatgaatatatacaaatgcttttgaaaaactcaaggaa	3588
Db	2341	CPTCAAGATTGGGAAGTTTCAGCCATGAATATACAAATGCTTTTGGAAAACTCAAGGAA	2400
Qy	3589	ttccccagttagaagccagagacatgataactgtgatacttcagatgtttgacacag	3648
Db	2401	TTCCCCAGTTTGAAGCCAGAGGACATGATAACGTGGGATCTTCAGATGTTTGAACACAG	2460
Qy	3649	tgaagcgttagaagaaaaattgtttaattgttagcaaccacatcaggatcggaattcta	3708
Db	2461	TGAAGCGGTTAGAGAAAAATCTTGTTAATTGTAGCAACACATCAGGATCGGAATCTA	2520
Qy	3709	ttaagaatgatgatattactcatgataaag	3738
Db	2521	TTAAGAATGATGAGATTACTCATGATAAG	2550
RESULT	5		
AR030785			
LOCUS	AR030785	1001 bp	DNA
DEFINITION	Sequence 2 from patent US 5861312.		
ACCESSION	AR030785		
VERSION	AR030785.1	GI:5943999	
KEYWORDS	Unknown.		
SOURCE			

ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 1001)		
AUTHORS	Varshavsky, A. and Kwon, Y. Tae.		
TITLE	Nucleic acid encoding mammalian UBRI		
JOURNAL	Patent: US 5861312-A 2 19-JAN-1999;		
FEATURES	Location/Qualifiers		
source	1..1001		
BASE COUNT	363 a 186 c 205 g 247 t		
ORIGIN	/organism="unknown"		
Query Match	15.8%; Score 996.2; DB 6; Length 1001;		
Best Local Similarity	99.7%; Pred. No. 2.6e-240;		
Matches 998; Conservative	0; Mismatches 3; Indels 0; Gaps 0;		
Qy	2799	atggatcccaataaagttctgttacttcaggttacttcagaggtatgaacttcgagggctttt	2858
Db	1	ATGGATCCCAACAAGTCTTTGTTACTTGGTACTTCAGAGGTATGAACCTTGCGAGGCTTTT	60
Qy	2859	acaagaccatactacaaaagaccaggtttgattaaacaataatacacactaataagaa	2918
Db	61	ARCAAGCCATACTACAAAAGACCAGGATTTGATTAACAATAATAACACTAATAGAA	120
Qy	2919	gaaatgcttcaggtcctcctatctatgttgggtgagcgttatgtacctggagtggaagt	2978
Db	121	GAATGCTTCAGTCCCTCATCTATTTGTGGTGGAGGTTATGTACCTGGAGTGGGAAT	180
Qy	2979	gtgaccaagaagaggtgcacaatgagagaataatcattcacttctgttgcattgaacccatg	3038
Db	181	GTGACCAAGAAGAGGTCACAATGAGAGAAATCATTCACTTGCTTTGCAATGAAACCATG	240
Qy	3039	ccacacagtcgcaatgcacaaaatttacctgagaatgaagaataatgaaactggcttagag	3098
Db	241	CCACACAGTGCCTATGTCACAAAATTTTACCTGAGAAATGAATAATGAAACTGGCTTAGAG	300
Qy	3099	aatgtcataaacaagtgccacatttaagaacaccaggtgtatcagggccatggagtttat	3158
Db	301	AATGTCTAAACAAGTGGCCACATTTAAGAAACCCAGGTGTATCAGGCCATGGAGTTTAT	360
Qy	3159	gaactaaagatgaatcactgaaagacctcaatatgtacttttacttactcctcaaaacc	3218
Db	361	GAATAAAAGATGAATCACTGAAAGACTTCAATATGTACTTTTATCATTTACTCCANAACC	420
Qy	3219	cagcatagcaagctgaacatatgcagaagaagaagagaaacaaagaaacaaagatgaa	3278
Db	421	CAGCATAGCAAGCTGAACATATGCAGAAAGAGAGAAACAAAGAAACAAAGATGAA	480
Qy	3279	gcattgcgcccacacacacctcctgaaattctgcctgcttctcagcaaaagtgttaacctt	3338
Db	481	GCATTGCGGCCACACACCCTCTGAAATTTCTGCCCTGCTTTTCAGCAAAGTGATTAACTT	540
Qy	3339	ctcaactgtgatcatcatgatgtacattctcaggaacgtatttgagcggggaatagacaca	3398
Db	541	CTCAACTGTGATATCATGATGTACTTCTCAGGACCGCTATTTGAGCGGGCAATTAACACA	600
Qy	3399	gattctaaactgtgaccgaaggtgctccaaatggcttttccatattcttggcattgggt	3458
Db	601	GATTCTAACTTTGGACCCGAAGGATGCTCCAATGGCTTTTCATATTTCTGGCAATGGGT	660
Qy	3459	ttactagaagagaagcaacagcttcaaaaagctcctgaagaagaagtaacatttgactttt	3518
Db	661	TTACTAGAGAGAAAGCAACAGCTTCAAAAAGCTCCTCAAGNAGAAAGTAACATTTGACTTT	720
Qy	3519	tatcataaggtctcaagattgggaagttcagccatgaataacataatgcttttggaaaaa	3578
Db	721	TATCATAAGGCTTCAAGATTGGGAAGTTCAGCCATGAATATACAAATGCTTTTGGAAAA	780
Qy	3579	ctcaaaaggaattccccagtttagaagccagagacatgataacgttgatacttcagatg	3638
Db	781	CTCAAAAGGAATCCCCAGTTAGAAAGGCCAGAGGACATGATAACGTGGATACTTCTCAGATG	840

Qy	3639	tttgacacagtgaaagcgttaagagagaaaaatcttgtttaattgtacgacacacatcagga	3698			
Db	841	TTTGACACAGTGAAGCGATTAAAGAGAAAAATCTGTGTTAAATGTAGCAACCATCAGGA	900			
Qy	3699	tcggaatctattaagaatgatgagattactcatgataaagaaaaagcagaacgaaaaaga	3758			
Db	901	TCGGAACTCTATTAAGAAATGATGAGATTACTCATGATAAAGAAAGACAGACGAAGAAGA	960			
Qy	3759	aaagctgaagctgcttaggtacatgccgagaagatcatggc	3799			
Db	961	AAAGCTGAAGCTGCTAGGCTTCATGCCAGAAAGATCATGGC	1001			
RESULT 6						
AR121464	LOCUS	AR121464 1001 bp DNA	PAT 16-MAY-2001			
DEFINITION	Sequence 2 from patent US 6159732.					
ACCESSION	AR121464					
VERSION	AR121464.1 GI:14105040					
KEYWORDS						
SOURCE	Unknown.					
ORGANISM	Unknown.					
REFERENCE	Unclassified.					
AUTHORS	1 (bases 1 to 1001)					
TITLE	Varshavsky, A. and Kwon, Y. Tae.					
JOURNAL	Nucleic acid encoding mammalian Ubr1					
FEATURES	Patent: US 6159732-A 2 12-DEC-2000;					
source	Location/Qualifiers					
	1..1001					
BASE COUNT	/organism="unknown"					
	363 a	186 c	205 g			
ORIGIN	247 t					
Query Match 15.8%; Score 996.2; DB 6; Length 1001;						
Best Local Similarity 99.7%; Pred. No. 2.6e-240;						
Matches 998; Conservative 0; Mismatches 3; Indels 0; Gaps						
Qy	2799	atggtacccaataagttcttgttacttggtacttcagaggtatgaactgccagagctttt	2858			
Db	1	ATGGATCCCAACAAGTCTTGTTACTGTGTTCTCAGAGGTATGAAC TTGCCGAGGCTTTT	60			
Qy	2859	aacaagaccatctctacaaagaccagattgtattaaacaataatacactaatagaa	2918			
Db	61	AACAAGACCATTATCTACAAAGACCAGGATTTGATTAAACAATAATAACATAAGAA	120			
Qy	2919	gaactctcaggtcctcatctatatttggtgaagcgttatgacctgaagtggaaat	2978			
Db	121	GAATGCTTCAGTGCTCTATCTATATGTGGGTGAGCCTTATGACCTGGAGTGGGAAT	180			
Qy	2979	gtgaccacaaagagaggtgcacatgagagaaatcattcacttgcttgcattgaacctg	3038			
Db	181	GTGACCAAGAAGAGGTCACATGATGAGAAATCATTACTTGC TTGCATTGAACCATG	240			
Qy	3039	ccacacagtgccattgccccaaaaatttaacctgagaatgaaataatggaactggcttag	3098			
Db	241	CCACACAGTGCCATTGCCAAAAATTTACCTGAGAATGAAAAATAATGAACTGGCTTAG	300			
Qy	3099	aattgtcataacaaagtgccacatttaagaacagggtgatcaggccatggaatttat	3158			
Db	301	AATGTCTATAACAAAGTGCCACATTTAAGAAACAGGTGTATCAGGCCATGGAGTTTAT	360			
Qy	3159	gaactaaagatgaatcactgaaagacttcaatgtactttttatctattactcctcaaac	3218			
Db	361	GAACTAAAGATGAATCACTGAAAGACTTCAATATGTACTTTTATCATTTACTCAAACC	420			
Qy	3219	cagcatagcaaggtgaaacatgacagaagaaaggagaaacaaagaaaaaagaatgaa	3278			
Db	421	CAGCATAGCAAGGCTGAACATATGCAGAGAAGGAGAAACAAAGAAACAAGATGAA	480			
Qy	3279	ggaattgccaccaccactcctgaaattctgcctgcttcttcagcaagtgattaacctt	3338			
Db	481	GGATTCGCCCCACACACACTCCTGAAATTCGCTGCTGCTTTACGAAAGTGATTAACTT	540			

Qy	3339	ctcaactgtgatacatgatgtacattctcaggaccgtatttgagcgggcaatagacaca	3398
Db	541	CTCAACTGTGATATCATGATACATTTCTCAGGACCGTATTTTGAGCGGGCAATAACACACA	600
Qy	3399	gattctaacttgtggaccgaaggatgcctcaaatggctttcttcatacttcggcattgggt	3458
Db	601	GATTCTAACTTGTGGACCGAAGGATGCTCCAAATGGCTTTTCATATTTCTGGCATTGGGT	660
Qy	3459	ttactagaagaagcaacacgcttcaaaagctcctctgaagaagaagtaacatttgacctt	3518
Db	661	TTACTAGAAGAAGAAGCAACACGTTCTCAAAAGCTCCTGAAGAAGAAGTAACATTTGACTTT	720
Qy	3519	tatcataagccttcaaatgtggagttcgacgatgaatatacaaatgccttttggaaaaa	3578
Db	721	TATCATRAGGCTTCAAGATTGGGAAGTTCAGCCATGAATATACAAATGCTTTTGGAAAAA	780
Qy	3579	ctcaaaagaattccccagtttagaaggcagaaagacatgataacgttgatacttcagatg	3638
Db	781	CTCAAGAGAAATTTCCAGTTTAGAAGGCCAGAAAGACATGATAACGTGATATTTCAGATG	840
Qy	3639	tttgacacagtgaagcgattaaagagaaaaatccttggtttaattgtagcaaacacatcagga	3698
Db	841	TTTGACACAGTGAAGCGATTAAAGAGAAAAATCTTGTTTAATTGTGAGCAACCATCAGGA	900
Qy	3699	tcggaattcttaagaatgatgaattacttactcatgataaaagaaaaagcagaacaaaaaga	3758
Db	901	TCGGAATCTATTAGAATGATGAGATTACATCATATAAGAAAAAGCAGAACGAAAAAGA	960
Qy	3759	aaagctgaagctgtaggtcacatcgcaagaagatcatggc	3799
Db	961	AAAGCTGAAGCTGTAGGCTTCATCGCCAGAAATCATGGC	1001
RESULT	7		
AF061556			
LOCUS	AF061556	999 bp	mRNA
DEFINITION	Homo sapiens ubiquitin-protein ligase E3-alpha (UBR1) mRNA, partial cds.		10-JUL-1998
ACCESSION	AF061556		
VERSION	AF061556.1	GI:3170888	
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 999) Kwon,Y.T., Reiss,Y., Fried,V.A., Herskho,A., Yoon,J.K., Gonda,D.K., Sangan,P., Copeland,N.G., Jenkins,N.A. and Varshavsky,A.		
TITLE	The mouse and human genes encoding the recognition component of the N-end rule pathway		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 95 (14), 7898-7903 (1998)		
MEDLINE	98318583		
REFERENCE	2 (bases 1 to 999)		
AUTHORS	Kwon,Y.T. and Varshavsky,A.		
TITLE	Direct Submission		
JOURNAL	Submitted (24-APR-1998) Division of Biology, 147-75, California Institute of Technology, 1200 E. California Bl., Pasadena, CA 91125, USA		
FEATURES			
source	Location/Qualifiers		
	1..999		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/map="15q15-q21.1"		
	/chromosome="15"		
	1..>999		
	/gene="UBR1"		
	1..>999		
	/gene="UBR1"		
	/note="UBR1p; recognition component of the N-end rule pathway"		
	/codon_start=1		
	/product="ubiquitin-protein ligase E3-alpha"		
gene			
CDS			

[illegible]

```

RESULT      8
AB002347    6158 bp      mRNA      PRI      13-FEB-1999
LOCUS       Human mRNA for KIAA0349 gene, partial cds.
DEFINITION  AB002347
ACCESSION   AB002347.1 GI:22244638
VERSION     AB002347.1
KEYWORDS    Homo sapiens male brain cDNA to mRNA, clone_lib:pBluescriptII SK
SOURCE      plus clone:HG1561.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 6158)
AUTHORS     Nagase,T., Ishikawa,K., Seki,N., Nakajima,D., Ohira,M.,
            Miyajima,N., Kotani,H., Nomura,N. and Ohara,O.
TITLE       Direct Submission
JOURNAL     Submitted (28-MAR-1997) to the DDBJ/EMBL/GenBank databases. Nobuo
            Nomura, Kazusa DNA Research Institute, Gene Structure 1; 1532-3
            Yana, Kisarazu, Chiba 292, Japan (E-mail:cdnainfo@kazusa.or.jp,
            URL:http://www.kazusa.or.jp, Tel:+81-438-52-3930,
            Fax:+81-438-52-3931)
FEATURES             2 (sites)
            Nagase,T., Ishikawa,K., Nakajima,D., Ohira,M., Seki,N.,
            Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
            Prediction of the coding sequences of unidentified human genes.
            VII. The complete sequences of 100 new cDNA clones from brain which
            can code for large proteins in vitro
            DNA Res. 4 (2), 141-150 (1997)
            97349984
            Location/Qualifiers
            1..6158
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone="HG1561"
               /clone_lib="pBluescriptII SK plus"
               /sex="male"
               /tissue_type="brain"
            1..3828
               /gene="KIAA0349"
               /c1..3828
               /gene="KIAA0349"
               /codon_start=1
               /protein_id="BAA20806.1"
               /db_xref="GI:22244639"
               /translation="LKVLISKPTENSDLRKQLEGFDAFLLELLKCMQGMDDPIITQV
               GEHIEPEWAEATLOMKLTHVISMQDWCSDEKVLGLIAYKKCLAVLLMCMQGGHYTD
               GQPTITLSCGHSVETIRYCVSQEKYSIHLPVLRLLAGLVLLSKSEVAYKFFELLPJL
               SELSPMLIEHRLPCLVCAQVHAGMRRNGFSLVQIYYHNKVKCRREMFDFVLMVL
               QTVGSMDDPHLMIMLSRFELQYIFSTPDYGRFSEITHKDVQQNNTLIEMLYL
               IIMLVGRFSPGVQVNADEIKREIHLQISIKPMASELVKSLPDEKNETGMSVI
               EAVAFKPGLTGRGMTELPEKAEKFNLYFIHPSRAEQSKAEAEQKLRKQNRDTA
               LPPVLPFPCLFASNLVQLQSDVLMICIMGTILOAWAEHNGYAWSSSMQKLRQNLHIGM
               ALQEKHGLNVEEHWVPTFTQIKISGPEAKNPISLAMILLETQNALPYLVDHKDM
               IRWLTKTFNAVKKMRSSPTSPVAETEGTIEEESRDKDAERKRAEIALRREKIM
               AQSEMRHFIIDENKELFOOTLEDAFSTVSLDHPSPVASTMTLTALGPTOTQYPEQRQ
               FVTCILQEQEYKVESRANVLAAVFQVRSTVLSKNRSKFIQDPEKVDPLPMHDPDLSG
               THYSSCGHIMHACWQRYFQVQAEQRQRRRLHSTSYDVENGEFCLPCCECLSTV
               IPLLPPRFINNRNLNSDPQNTQWITLSQIKALQFLRKKEESTPNNAATKNSENV
               DEQLQPFGRFDPKPIFYGESIKEMLTFTGTATYKVLKVFHEEDPRVPIKMGSCG
               AYTQTSRTERILSDKPLFGPLRCLPDRKSTPFAAAHWTVAASVVGQGHFKCLFA
               SLVNDSESLPCLDIDIMEHLLGLVLAFLPQACDFSGISLGTGLHFLHFTVAWIA

```


IIQILLTSCTEENGMDENPPCEESAVILALYKTLHOYTGSALKETPSGWHLSVRA
GIMPELKCSALFHFVLYNGVPSPPDIQVPGTSHFEHLCSYLSLNNLICFQENSETM
SLIESWCRNSEVRILEGERDAIRPRESKINLINPEDYSSLINQASNFSCPKSGGDK
SRAPTCLIVCGSLICSQSYCCQTELEGGACTAHTYSCGSGVGIFLVRRECQVFL
AGTKGCFYSPYLDYGETDQGLRRGNPLHLCKERFKKIQLKWHQSHVTEIGHAE
ANOTLVGIDWQHL"

BASE COUNT 1832 a 1199 c 1320 g 1807 t
ORIGIN

Query Match 15.6%; Score 981; DB 9; Length 6158;
Best Local Similarity 55.9%; Pred. No. 2.1e-236;
Matches 2157; Conservative 0; Mismatches 1610; Indels 90; Gaps 12;

QY	2130	ctaaagatatctgcgtacgcagcaaccacaataatgacagagaagattaaagaatgcagttc	2189
DB	1	CTCAAGTATGCTTAATAGCAACCAACTGAATGGTCAGATGAGCTGAGGCAGAAAGTTC	60
QY	2190	cittgaaggttttcgatctcttttttgaagattcttacctgtatgcaggggaatggagaataatc	2249
DB	61	CTAGAAAGGGTTTGATCCCTTTTGGAAATTACTAAATGTATGACGGGAATGGATCCAATT	120
QY	2250	cgaagacaggttgggcaacacattgaagtggtccctgattggaggtcgccattgctata	2309
DB	121	ACACGTCAGTAGGACAACTATTGGAATGGACACAGATGGGAAGCAGCCTTCAACACTA	180
QY	2310	cagatgcaattgaagaatatcttactcatgttccaagagtggtgtcctgtgatgaaga	2369
DB	181	CAAAATGAATTAACACATGTCATTTCATATGATGCAGGACTGGTGCTTCAGATGAAAA	240
QY	2370	ctcttacttggtgcttataaagaatgcacaaagctgtagtgggtgcagtaaccagtttc	2429
DB	241	GTGTTAATCGAAGCTTTACAAAGAAATGCTCGCTGTACTGATGCAGTGTCTCATGGTGTAT	300
QY	2430	atatctagtagcaagacagt--agfacaatcgttggacatagtttggaaacaaagttcc	2486
DB	301	ACTGATGGTGACAGGCCATCACACTTAGCAATTTGGACATTCAGTGGAAACTATCAGA	360
QY	2487	tacagagtatctgagagatcttgaagacatacatctgccactctctcaggacccttgcgtgt	2546
DB	361	TACTGTGTTTCCCAAGAAAAAGTTAGCATTCACCTCCACGTTTCTCGCTTACTTGCAGGT	420
QY	2547	ccttcagtacgtttaagacaggtcgggtgcgtgttctcaagactgcatagaatttgctctttt	2606
DB	421	TTACATGTATTATTAAAGCAAAAGTGAAGTGGCATATAAATTTCCAGAGCTCCTACCTCTA	480
QY	2607	gaggactttcaagtagagtagtactagtggaatctccttaactgttgcgtgtgttggttgc	2666
DB	481	AGTGAACTTAGCCACCCCATGTTGATAGAACACCCCTCTTAGATGTCTTGTCTGTGTGCC	540
QY	2667	cagggttctgctgagatgtggcgaagaataatggactgtctcttattagccaggtgttttat	2726
DB	541	CAAGTACATGCCGAATGTGGAGAAGAATGGTCTCTCTAGTAACACAGATTTATTAC	600
QY	2727	taccagaatgtttaagtcgagagaagaataatgtatgataaagatatcatctgccttcagatt	2786
DB	601	TACCATAATGTGAAATGCAGACGTGAGATGTTTGACAAAGGATAGTAAATGCTTCAGACA	660
QY	2787	ggtgcattttaatggatcccaataagttcttgttactggtacttcagaggtatgaactt	2846
DB	661	GGTGTCTCCATGATGCCAATCAATTTCCCTGATGATCATGTCTAGCCGCTTTGAACTT	720
QY	2847	gccgaggtctttaacaagaccatatctacaaa-----gaccag	2885
DB	721	TATCAGATTTTTCAGTACTCCAGACTATGGAAGAAAGATTTAGTTCTGAGATTACCCATAAG	780
QY	2886	gatttgattaaacaataataacactaataagagaataatgcttcagggtccctcatctatt	2945
DB	781	GATGTGTTCAGCAGAACAACTACTCTAATAGAGAATGCTATACCTCATTAATAGCTTT	840
QY	2946	gtgggtgagcgttatgtacctggagtgggaatgtgacccaaagaaggttcacaaatgaga	3005
DB	841	GTTGGAGAGAGATTTAGTCTCTGGAGTTGGACAGGTAAATGCTACAGATGAAATCAAGCGA	900

QY	3006	gaatcattcactctgtcttgcattgaaccccatgcccacacagtgccattgcccacaaattta	3065
DB	901	GAGATTATCATCATGTTGAGTATCAAGCCTATGCTCATAGTGAATTGGTAAAGTCTTTA	960
QY	3066	cctgagaatgaaaaataatgaactgcttagagaagtgcataaacaagaatggccacattt	3125
DB	961	CCTGAAGATGAGAAAGGAGACTGGCATGGAGATGTAATCGAAGCATTTGCCCATTTTC	1020
QY	3126	aagaaccaggtgtatcagggccatggagtttatgaactaaagaatgaatcactcagaagac	3185
DB	1021	AAGAAACCTGGATTAAACAGGACGAGGCATGTATGAATCTGAAACCAAGATGTGCCAAGAG	1080
QY	3186	ttcaatgtacttttattctccaaatctccaaacccagcagtagcaaggtggaacatatgcag	3245
DB	1081	TTCAACTGTGATTTCTATCACTTTTCAAGGGCAGAACAGTCCAAGGCAGAGAAGCGCAA	1140
QY	3246	aagaaaaaggagaaaaaacaagaacaaagatgaagcattgcgcgcacccaccctcctgaa	3305
DB	1141	CGGAAATTGAAAGACAAAAATAGAGAAGATACAGCACTCCCACTCCGCTGTTGGCTCCA	1200
QY	3306	ttctgcctcgtctttcagcaaatgatttaacctctcaactgtgatatcatgatgatcatt	3365
DB	1201	TTCTGCCCTCTGTTTGCNAGCCTGGTTAAACATTTTTCAGTTCAGATGTCATGTTGTGCATC	1260
QY	3366	ctcaggaccgtatttgagcgggcaatagacacagattcttaaccttgggacgcgaaggatg	3425
DB	1261	ATGGGAACAATTCCTGCAATGGCTGTGGAACATAATGGATATGCCCTGGTCAGAGTCCATG	1320
QY	3426	ctccaaatggcttttctaatctctggcattgggttttactagaagaagaacaacagctcaa	3485
DB	1321	CTGCAAAAGGGTGTACATTTAATTTGGCATGGCCTACAGCACTACAAAGAAAGAAACAACTTAGAG	1380
QY	3486	aaagctcctgaag--aagaagtaacatttgactttttatcataagccttcaagattggga	3542
DB	1381	AATGTCAGGAGAGACATGTAGTACATTTACCTTCACTCAGAAATATCAAAACCTGCT	1440
QY	3543	agttcagccatgaat-----atacaaatgcttttggaaaaactcaaaaggaattccc	3593
DB	1441	GAAGCGCCAAAAAATTCCTCTAGCATACTAGTATGCTGGAAACACTACAAAAATGCTCCC	1500
QY	3594	cagttagaagccagaagagacatgataaactggatctcagatcttggacacagtggaag	3653
DB	1501	TACTAGAAGTCCCAAAAGACATGATTCGGTGGATATTGAAAGACTTTTAATGCTGTAAA	1560
QY	3654	cgattagaagaaaaatcttgtttaattgttagcaaccacatcagatcggaatcttattag	3713
DB	1561	AAGATGAGGAGA-----GTTCACTACCAAGTCCCGTGGCAGACAGAGAAGAACCATTA	1614
QY	3714	aatgatgagattactcatataaagaaaaagcagaacgaaaaagaaagctgaaagctgct	3773
DB	1615	ATGGAAGAGAGTTCAAGGGCAACAAAGACAAAGCTGAGAGGAAGAAAGACAGAGATTGCC	1674
QY	3774	aggctacatgcgcagaagatcatggtcagatgtctgcttacagaaaaaacttcattgaa	3833
DB	1675	AGACTGGCAGAGAAAAAGATCATGGCTCAGATGCTGAAATGTCAGCGGCATTTTATTGAT	1734
QY	3834	actcaataactcatgtatgacaatacatcagaataatgcctgggaagaagattccattatg	3893
DB	1735	GAAACAAAGAAACTCTTTTCAGCAGACATTAGAACTGGATGCTCAACCTCTGCTGTTCTT	1794
QY	3894	gaggaagagagcaccgccagtcagtgactactcttagaattgttttgggtcctaaacgg	3953
DB	1795	GATCA-----TAGCCCTGTGGCTTCAGATATGACACTTACAGCACTGGGTGCCACACAA	1848
QY	3954	ggttcactgttactgaaaaggaggtgctgacgtgacatccttttggcaagaagaacagag	4013
DB	1849	ACTCAGGTTCCTGAAACAAAGACAATTCGTTACATGTAATTGTTGTCGAAGAGGAGCAAGAA	1908
QY	4014	gtgaaataagaaaaataatgcccattggttatctcgccgtgtgtccagaataatctactgctta	4073
DB	1909	GTTAAAGTGGAAAGCAGGGCAATGGTCTTGGCAGCAATTTGTCAGAGATCAACTGTATTA	1968

Qy	4074	accgcagcacagggaagaaacccaataagaactctcaggagaagccctagaccocaatttccatg	4133
Db	1969	TCAAAAAACAAGATAAATTTATTCAAGATCCAGAAAAATAT---GATCCATTATTTCATG	2025
Qy	4134	gatccagacttgccatataggaaactatatacaggagaagctgtggttcattgtaataagcagcagtg	4193
Db	2026	CACCCTGATCTGCTGTGTGGAAACACACACTAGTAGCTGTGGGCACATATTATGATGCCCAT	2085
Qy	4194	tqctggcagaagatattttagaagctgacagctgagctctcagcagcgcacattcattgtgac	4253
Db	2086	TGTTGGCAAAAGGTATTTTGATTTCCGTTCAAGCTAAAGAACACGGGAGGCAACAGAGATTTA	2145
Qy	4254	ctttt-----tgacttgaaagtgagaaatatctttgccctctcttgcaaatct	4301
Db	2146	CGCTTACATACGAGCTATGATGTAGAAAACGGAGAATTCCTTTGCCCTCTTGTGGAATGC	2205
Qy	4302	ctgtgcaataactgtatccccatttctcttgcaacctcaaaaagataaaacagtgagaat	4361
Db	2206	TTGAGTAATACTGTTATTATCC-----TCTGCTGCTCTCCCAAGAAATATTTTAAACAAC	2259
Qy	4362	gcagatgctcttgctcaactttgacctggcacggtgatacacagacttctctggccaga	4421
Db	2260	AGGTTAAATTTTTCAGACCAACCAATCTGACTCAGTGGATTTAGAACAAATATCTCAGCAA	2319
Qy	4422	atatcaggttataataagaacatgctaaaggagaaaacccaattctatttctttaa	4481
Db	2320	ATAAAGGACATTACGTTTCTTAGGAAAGAAAGAAAGTACTCTCTAAATAATGCCCTACAAAG	2379
Qy	4482	caaggaaatgggagattctactttggagttccattccatctctgagttttggcgttgagct	4541
Db	2380	AATTTCAGAAAATGTGGATGAATTCAGAGTCCCTCGAAGGGTTTCAGGGCTGATTTTCGTCT	2439
Qy	4542	tcgatataatttcaaaatagcatcaagaaatgggttatctcttctggccacaacaattat	4601
Db	2440	AAGATCCCTTATTCTGAGAGCATAAAGAAATGCTAACGACATTTGGAACACTGCTACTAC	2499
Qy	4602	agaattgattgaagtgcaactgataaaggatctctcagtcgccactgctgcaoctgg	4661
Db	2500	AAGGTGGGACTAAAGGTTTCATCCCAATGAAGAGGATCTCGTGTCCCATTAATGTGTGG	2559
Qy	4662	agcaactgcgcttccactatccaggcaacttgaaaatctatcttgggagatgaagaaaaacct	4721
Db	2560	GGTAGCTCGCGCTACACCATCCAAAGCATAGAAAGAAATTTTCAGTGATGAAGATAACCA	2619
Qy	4722	ctgtttgagcaacttcaaaataggcagcataatgctctgaagcacttaatgcagtttgc	4781
Db	2620	TGTTTTGGTCCCTTACCTTGCAGACTGGATGACTGCTTTAGGTCAATGAGGATTTGGCC	2679
Qy	4782	gttcacagagattacctgtcctcaggtcctgatacagaacaactggtctgctctcta	4841
Db	2680	GCAGCACACTGGACAGTGSATCAGTTTCAGTGGTGCAAGGACATTTTGTAAACTTTT	2739
Qy	4842	tcagttgtcttcttaacataaaataagaagatacaccaatgcctctctgtctatagatctg	4901
Db	2740	GCATCACTGGTGCCCTAATGACAGCCATGAGAACTTCCATGCATATATAGATATTTGACATG	2799
Qy	4902	tttcatgttttgggtggctgtgttagcatctccatctctgtattgggatgaccctgtt	4961
Db	2800	TTTCATTTATTGCGTGGGCTTGGTGTGCTATTTCCGCGTTCGAGTG-----2846	
Qy	4962	gatctgcagcctcttccagttagttcttctataacaccttattctcttccatttgatc	5021
Db	2847	--TCAGGATTTTTCAGGGATCAGCCTTGGCACTGGAGACCTTCACATTTTCCATCTGGTT	2904
Qy	5022	accatggcacacatgcttcagatactaacttaacagttagacacagccc---taccacctgtct	5078
Db	2905	ACTATGGCACACATCATACAGATCTTACTTACCTTACCTCATGTACAGAGAGAAATGCGATG	2964
Qy	5079	caggttcaagaagacagtgaagaggtcattccgcatctcttcttcttgcagaaattct	5138
Db	2965	CAAGAAATCCCCCTTGTGAAGAAGAAATCAGCAGTTTCCTTGTGTATAAACACTTTCAC	3024
Qy	5139	caatatacaagtggtccatttgggtgtgatatctctggctggctattgttgggtctcactg	5198

Db	3025	CAGTATACGGGAAGTGCCTTTGAAAGAAATACCATCGGCTGGCATCTGTGGAGGAGTGTC	3084
Qy	5199	aagaatgcaatcacacccttatctcgctgtgtgcattgtttttccactatttacttggg	5258
Db	3085	AGAGCTGGAATCATGCTTTCTGGAAGTGTTCGCTTTATTTTCTATTAATAATGGA	3144
Qy	5259	gtaactccgcctgaggaactgcataccaattctgcagaagagagtagacagtgcactctgt	5318
Db	3145	GTTCTCTCCCAACCCGACATTC--AGTTCCTGGAAACAGCCATTTTGAACATTTATGT	3201
Qy	5319	agctatctatcttctaactacaattgttctctgtctcttccagggaataatttgggatactgta	5378
Db	3202	AGCTATCTTTCCTACCAACAACCTCATTTGCTCTTTTCAAGAAAATAGTGAGATAATG	3261
Qy	5379	agggccctgtctccagagggtgggtgcagatcctgccttacttaaacgtgttgaagcaaaaa	5438
Db	3262	AATTCACTCATGTTGAAAGTTGGTCCGTACAGTGAAGTTAAAAGATATCTAGAAAGTGAA	3321
Qy	5439	aaacacgtgtgcaggtacccttagaanaagaaatagttgatagagcttctcgtgatgactat	5498
Db	3322	AGAGATGCTATAAGATATCCAAGAGAACTTAACAATTAATAACCTTCACAGGATTAC	3381
Qy	5499	agctgcctcctgaaataagcttctcatttcaggtgcccaagcgtctcgaatgatgagcga	5558
Db	3382	AGCAGCCTCATTAATCAAGCATCCAAATTTCTCGTCCCGAAATCAGGTGGTGATAAGAGC	3441
Qy	5559	aageatcctgtcctctgccttcttctgtggggtatactatgtctcagaacaatttgcctgc	5618
Db	3442	AGAGCCCCAACCTGTGTGCCCTGTGTGCGGATCTCTGCTGTGCCAGAGTTACTTGCTGC	3501
Qy	5619	caggaaattgaaagggaagaggttgagcttgcatcttttcacgcaactcactgtgga	5678
Db	3502	CAGACTGAATGGAAGGGGAGGATGTAGAGCCTGCACAGCTCACACCTACTCTGTGGC	3561
Qy	5679	gccggagctgcgatttcttcctaaaaatcagagaaatccgagtgccgagtggtcctggttgaaggtaaa	5738
Db	3562	TCGGAGTGGGCATCTCTCGAGAGTACGGGAATGTCAAGTGCTATTTTAGCTGGCAAA	3621
Qy	5739	gccagagctgtgcctatccagctccttacttgatgaataatagagaaacagaccctggc	5798
Db	3622	ACCAAAGGCTGTTTTTATTTCTCTCTTACCTTGATGACTGTGGGAGACCCGACAGGA	3681
Qy	5799	ctgaagaggggcaacccccttcatttatctcgtgagcgggtatcggaagctccatttggct	5858
Db	3682	CTCAGACGGGGAATCCTTTACATTTATGCAAGAGCGGATTCAGAGAGATTCAGAAGCTC	3741
Qy	5859	tggcaacaacatgcattatagaagagattgtctagagccaaagagactaatcagatgta	5918
Db	3742	TGGCACCAACACAGTGTACAGAGGAAATTTGGACATGCACAGGAAGCAATCGACACACTG	3801
Qy	5919	tttggattcaactggca 5935	
Db	3802	GTTGGCATTTGACTGGCA 3818	
RESULT	9		
HSX521			
LOCUS	HSX521	818 bp	DNA
DEFINITION	Human partial cDNA sequence, clone x521; .		PRI
ACCESSION	247040		
VERSION	247040.1	GI:1008975	
KEYWORDS	human.		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomom		
REFERENCE	1 (bases 1 to 818)		
AUTHORS	Chhannilkulchai, N., Pasturaud, P., Richard, I., Auffray, C. and Beckmann, J.S.		
TITLE	cDNA selection in the LGMD2A region		
JOURNAL	unpublished		
REFERENCE	2 (bases 1 to 818)		

Db	663	CAAGACATCAAAATGACGACAGTAGTAAGGCTGATTCACAAATATATGAAAACTTCT	722
Qy	6021	gagqctgggaaagtattgagggctcttttgctccatgtccaggttcacttacatcaata	6080
Db	723	GAGGGCTGGTGAAGTATTGAGGGCTCTTTTGCTCCATGTCCAGGTTCACTTACATCAATA	782
Qy	6081	aaatattcttaatgga	6097
Db	783	AAATATTCTTAATGGA	799
LOCUS	AK026998	2958 bp	29-SEP-2000
DEFINITION	Homo sapiens	CDNA: FL23345	fis, clone HEP13679.
ACCESSION	AK026998		
VERSION	AK026998.1	GI:10439998	
KEYWORDS	oligo capping; fis (full insert sequence)		
SOURCE	Homo sapiens	hepatoma cell_line:HepG2	CDNA to mRNA, clone_lib:HEP
ORGANISM	Homo sapiens		
REFERENCE	1 (sites)		
AUTHORS	Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y., Okamoto,S., Ohtani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.		
TITLE	NEDO human cDNA sequencing project		
JOURNAL	Unpublished (2000)		
REFERENCE	2 (bases 1 to 2958)		
AUTHORS	Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.		
TITLE	Direct Submission		
JOURNAL	Submitted (29-AUG-2000)		
COMMENT	to the DDBJ/EMBL/GenBank databases. Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center; Shiokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:cdnal@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)		
FEATURES	NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing: Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).		
source	Location/Qualifiers		
	1. .2958		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/cell_line="HepG2"		
	/cell_type="hepatoma"		
	/clone="HEP13679"		
	/clone_lib="HEP"		
	/notes="cloning vector pME18SFL3"		
BASE COUNT	940 a	600 c	638 g
ORIGIN	780 t		
Query Match	10.3%	Score	651.6; DB 9; Length 2958;
Best Local Similarity	54.6%	Pred. No.	2.8e-153;
Matches 1555; Conservative	0;	Mismatches 1229;	Indels 66; Gaps 1
Qy	3113	agtggccacatttaagaaaccaggtgtatcaggccatggagttatgaactaaagaatga	3172
Db	9	AGTTGCCCATTTCAAGAAACCTGGATTACAGACGACGAGCATGTATCAACTGAAACCCAGA	68
Qy	3173	atcactgaagactcaatgatgtactttatcttattctatctccaaaccagcatgacgaagc	3232
Db	69	ATGTGCCAAGAGTTCAACTTGTATTCTATCACTTTCAAGGGCAGACAGTCCAAAGC	128
Qy	3233	tgaacatatgcagaagaaaggagaaacaagaacaaagaatgaagcattgcgcacc	3292

Db	129	AGAAGAAGCGCAACGGAAATTGAAAGAACAATAATAGAGAAGATACACGACTCCCACTCC	188
Qy	3293	accacctcctgaattgcctgctttcagcaaaagtgattaaacctctcaactgtgat	3352
Db	189	GGTGTGCCCTCCATCTCTGCCCTCTGTTGGCAAGCCTGGTTAACATTTTGGCAGTCAGATGT	248
Qy	3353	catgatgtacattctcaggaaccgtatttgagcggcaatagacacagatttcaacttgg	3412
Db	249	CATGTTGTGCATCATGGGAACAATTCGAATGGCTGTGGAACATAATGGATATGCCCTG	308
Qy	3413	gaccgaaggagtgcctccaaatggcttttcattctggcatggggtttactagaagaagaa	3472
Db	309	GTCAAGAGTCATGCTGCAAAAGGCTGTACATTTAATTTGGCATGCGCACTACAAGAAGAAA	368
Qy	3473	gaacagctctcaaaagctccctgaag---aagaagtaacatttgaacttttcatcaaggc	3529
Db	369	ACAACATTTTAGAGATGTCAGGAGAGCATGTAGTAACATTTTACCTTTACTCAGAAGAT	428
Qy	3530	tccaagtgtggaaagtccagccatgaat-----atacaaatgcttttggaaaaaact	3580
Db	429	ATCAAAACCTGGTGAAGCGCCAAAAATTCCTTAGCATACTAGCTATGCTGGAACACAT	488
Qy	3581	caaaagaaattccccagtttagaaggccagaagaacatgataaacgtggatacttcagatgtt	3640
Db	489	ACAAAATGCTCCCTACCTAGAAAGTCCCAAAAGACATGATCGGTGGTTATTGAAGACTTT	548
Qy	3641	tgacacagtgaagcgattaaagaaaaatctgtttaatgtgcaaccacatcaggtac	3700
Db	549	TATGCTGTTTAAAGATGAGGGAGAGTTTCACTTACCNGTC-----CCGTGGCAGAGAC	602
Qy	3701	ggaaactattaagaatgatgattactcatgataaagaaaaagcagaacgaagaaagaa	3760
Db	603	TGAAGAACCATTAATGGAAGAGAGTCAAGGAGCAAAAGACAAAGCTGAGAGGAAGAGAAA	662
Qy	3761	agctgaagctgtaggctacatgccagaagaatcatggctcagatgctcgccttaacagaa	3820
Db	663	AGCAGAGATTGCCAGACTCGCAGAGAAAGAAATCATGGCTCAGATGCTGAAATGCAGCG	722
Qy	3821	aaacttcattgaaactcataaactcatgtagcaaatatcatcagaatgctcgtggaaaga	3880
Db	723	GCATTTTATTGATGAAMACAAAGAACTTTTCAGCAGACATTAGAACTGGATGCCCTCAAC	782
Qy	3881	agattccattatggaggaaagagacaccccgagctcagtgactactctagaattgcttt	3940
Db	783	CTCTGTGTTCTTGATCA-----TAGCCCTGTGGCTTCAGATATGACACTTACAGCACT	836
Qy	3941	gggtcctaaacggggtccatctgttacttgataaaaggaggctgcagctgcacaccttggcca	4000
Db	837	GGGCCCCACACAAACFCAGTGTTCCTGAAACAAAGACAATTCGTTACATGTATATTGTGTCA	896
Qy	4001	agagaacaggaggtgaaaaatagaaaaaLaatgccatggtattatcgccctgctgcagaa	4060
Db	897	AGAGGACCAAGATTAAAGTGAANGCAGGCAATGGCTTTGGCAGCAATTTGTTTCAGAG	956
Qy	4061	atctactgcttaaccagcacagggggaaaaaccatagaacctcaggaagacctcaga	4120
Db	957	ATCAAACTGATTATCAAAAAACAGAAGTAAATTTATTCAAGATCCAGAAAA---TATGA	1013
Qy	4121	cccactttcatggatccaagacttggcataatggaacttatacagaagaagctgtggctatgt	4180
Db	1014	TCCATTATTATGACCCCTGATCTGTTGTGGAAACACACACTAGTAGCTGTGGGCACAT	1073
Qy	4181	aatgcacgcagtgctgcgcagaagtatttgaagctgtacagctgagctcctcagcagc	4240
Db	1074	TATGCATGCCATTGTTGGCAAAAGGTATTTTGTATTCCTGTTCAAGCTTAAGNACACGAAG	1133
Qy	4241	cattcatgttgacctttt-----tgacttggaaagttggagaatatctttggccc	4288
Db	1134	GCAACAGATTACGCTTACATACGAGCTATGATGTAGAAAACGGAGAATTTCCTTTGCCCC	1193
Qy	4289	tctttgcaaatctctgtgcaatctactgtgatcccatatttctcttggcaacctcaaaagt	4348

QY 5426 tttaagcaaaaaaacaccggtggtcaggtaacctagaaaaagaataagtttgatagagct 5485
 Db 2310 TCTAGAAGGTGAAGAGAGATGCTATAAGATATCCAAGAGAAATCTAACAAATTAATAAACCT 2369
 QY 5486 tcttgatgactatagctcctccctgaatcaagctctcattcaggtggtccacggtctgc 5545
 Db 2370 TCCAGAGGATTACAGAGCCTCAATTAATCAAGCATCCAAATTCYCGTCCCGCAATCAGG 2429
 QY 5546 agatgatgagcaagaagcatcgtctcctctgctcttctgtgtgggctatactatgtctca 5605
 Db 2430 TGGTATAAGAGCAGAGAGCCCAACTCTGTGCTTGTGCGGATCTCTGCTGTGCTCCA 2489
 QY 5606 gaacatttgctccagagaattggaacggggaagaggttgagcttgctatttttcaagc 5665
 Db 2490 GAGTTACTGCTGCCAGACTGAACATGGAAGGGGAGGATGTAGGAGCCTGCACAGCTCACAC 2549
 QY 5666 acttcacttggaagcggagctgctcatttctcttaaaatcagaaatccgagtggtcct 5725
 Db 2550 CTACTCTGTGGCTCTGGAGTGGCATCTTCCTTATCTCCTTACCTTGATGACTATGGGGA 2609
 QY 5726 ggttgaaggttaagcagaggtgctgctatccagctccttacttggaatggaatggaga 5785
 Db 2610 TTTAGCTGGCAAAACCAAGCTGTTTTTATCTCCTTACCTTGATGACTATGGGGA 2669
 QY 5786 acagaccctggcctgaagaggggcaacccctcatttctcgtgagcggatcgga 5845
 Db 2670 GACCGACAGGGACTCAGAGGGGAAATCCTTTATCATTTATGCAAGAGCGATTCAAGAA 2729
 QY 5846 gctccatttgctggaacaaactgcattatagagagattgctaggagccaagagac 5905
 Db 2730 GATTCAGAGCTCTGGCACCACACAGTGTACAGAGAAATGGACATGCACAGGAAGC 2789
 QY 5906 taatcagatgttatttgattcaactggca 5935
 Db 2790 CAATCAGACACTGCTGGCATTTGACTGGCA 2819

RESULT 11
 AC090514/c
 LOCUS
 DEFINITION Homo sapiens chromosome 15 clone RP11-2016 map 15q15, complete sequence.
 AC090514
 AC090514.1 GI:13129421
 VERSION HTG.
 KEYWORDS human.
 SOURCE
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 166518)
 Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G., Kaur,A., Madan,A., Nesbitt,R., Traicoff,R. and Hood,L.
 Sequencing of human chromosome 15 D15S146-D15S117 region
 Unpublished
 2 (bases 1 to 166518)
 Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G., Kaur,A., Madan,A., Nesbitt,R., Traicoff,R. and Hood,L.
 Direct Submission
 Submitted (27-FEB-2001) Multimegabase Sequencing Center, Institute for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA 98105, USA
 3 (bases 1 to 166518)
 Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G., Kaur,A., Madan,A., Nesbitt,R., Traicoff,R. and Hood,L.
 Direct Submission
 Submitted (01-MAR-2001) Multimegabase Sequencing Center, Institute for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA 98105, USA

----- Genome Center

Center: Multimegabase Sequencing Center
 Center code: UMWSC
 Web site: http://chroma.mbt.washington.edu/msg_www
 Contact: leerowen@systemsbiology.org
 Drafting center: WIBR
 ----- Summary Statistics -----
 Sequencing vector: pUC18; L08752
 Chemistry: Dye-terminator Big Dye; 90% of reads
 Chemistry: Dye-primer Big Dye; 10% of reads
 Assembly program: Phrap; version 0.990399

FEATURES
 source
 1. 166518
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="15"
 /map="15q15"
 /clone="RP11-2016"
 /clone_lib="RPC1 human BAC library 11"
 /note="This clone overlaps RP11-90M11 AC068727 and RP11-473C18 AC068724"
 1. 11505
 /note="overlap with RP11-90M11 AC068727"
 47150..47190
 /note="low quality data"
 106560..106790
 /note="low quality data"
 109345..109360
 /note="low quality data"
 114180..114240
 /note="low quality data"
 156815..166518
 /note="overlap with RP11-473C18, AC068724"
 BASE COUNT 51718 a 34493 c 32351 g 47956 t
 ORIGIN

Query Match 8.0%; Score 503.8; DB 9; Length 166518;
 Best Local Similarity 99.6%; Pred. No. 8.4e-116;
 Matches 505; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5802 aagaggggcaaccccttcatttatctcgtgagcggtatcggaagctccatttggtctgg 5861
 Db 41822 AGGAGGGGCAACCCCTTCATTATCTCGTAGCGGTATCGAAGCTCCATTTTGGTCTGG 41763
 QY 5862 caacaacactgcattatagaagagattgctagagcgaagagactaatcagatttat 5921
 Db 41762 CAACAACACTGCTATTATAGAAGAGATTGCTAGGAGCCAGAGACTAATCAGATTGTTATT 41703
 QY 5922 ggattcaactggcagttactgtgagctccaactctgcctcaagacaatcacaaatgacga 5981
 Db 41702 GGATTCAACTGGCAGTTACTGTGAGCTCCAACCTCGCTCAACACAATCACAATGACGA 41643
 QY 5982 cagtagtaagagctgattcaaaaattatgaaaacttctgagggctgggaaaagtattgga 6041
 Db 41642 CAGTAGTAAGGCTGATTCAAAATTTATGAAAACCTTCTGAGGGCTGGGAAGATTATGGA 41583
 QY 6042 ggggttttgcctcattgctccaggttcactatcataataataattttcttaaggagat 6101
 Db 41582 GGGTCTTTTGGTCCATGTCAGGTTCCACTTACATCAATAAATAATTTCTTAAATGGAGTAT 41523
 QY 6102 tgccttcattagcaaacatagtcttcacagaaagacacatagatcaatctgttta 6161
 Db 41522 TCGTTTCAATTAGCAAAACATATGCTTCAGGAAAAAGACATAGATCAATCTGTTTTA 41463
 QY 6162 tctgctagtttccagggaatttattcccttcataatttctctcaatttcaatttttc 6221
 Db 41462 TGTGCTAGTATTTCAGGAATTTATTCCTTCATAATTTGTCTCATTTTCAATTTATTTC 41403
 QY 6222 atccacttgtagatgaagtcacgcgtcaaacagctgttagacattttatgtgtggttaact 6281
 Db 41402 ATCCACTTGGTAGTAGAGTCACGTCACACAGTGTGAGACATTTTATATGTTGTTAACT 41343

```
QY 6282 cttctgcaatttctgatttggtgtttt 6308
|||||
Db 41342 CTTCTGCCAATTTTGTATTTGGTGTTT 41316

RESULT 12
AK026948 2512 bp mRNA PRI 29-SEP-2000
LOCUS Homo sapiens cDNA: FLJ23295 fis, clone HEP10637.
DEFINITION AK026948
ACCESSION AK026948
VERSION AK026948.1 GI:10439928
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens hepatoma cell_line:HepG2 cDNA to mRNA, clone_lib:HEP
clone:HEP10637.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (sites)
AUTHORS Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
Shibahara,T., Tanaka,T. and Nakamura,Y.
NEDO human cDNA sequencing project
Unpublished (2000)
2 (bases 1 to 2512)
AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
Shibahara,T., Tanaka,T. and Nakamura,Y.
Direct Submission
Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio
Sugano, Institute of Medical Science, University of Tokyo,
Laboratory of Genome Structure Analysis, Human Genome Center;
Shirokane-dai, 4-6-1, Minato-Ku, Tokyo 108-8639, Japan
(E-mail:cdna@iems.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing: Research Association for Biotechnology; cDNA library
construction, 5'- and 3'-end one pass sequencing: Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).
FEATURES
source
1..2512
location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="HepG2"
/cell_type="hepatoma"
/clone="HEP10637"
/clone_lib="HEP"
/note="cloning vector pHE18SFL3"
BASE COUNT 785 a 511 c 543 g 673 t
ORIGIN

Query Match 7.7%; Score 487.8; DB 9; Length 2512;
Best Local Similarity 54.1%; Pred. No. 6.1e-112;
Matches 1220; Conservative 0; Mismatches 987; Indels 50; Gaps 9;

QY 3694 caggatcggaatcttaagaatgatgagattactcatgataaagaaagcagaacgaa 3753
|||||
Db 150 CAGAGCAGAAGAACCAATATGGGAAGAGAGTTCAAGGGACAAAGACAAAGCTGAGAGGA 209
|||||
QY 3754 aaagaaagctgaagctgctaggtacatcgccagagaagatcatggctcagatgctgcct 3813
|||||
Db 210 AGAGAAACACAGAGATGGCCAGACTCGCGAGAGAAAGATCATGGCTCAGATGCTGAAA 269
|||||
QY 3814 tacagaaaaacttcattgaaactcatataactcatgatatgacatacatcagaatgcctg 3873
|||||
Db 270 TGCAGGGCATTATTTATTTGATGAAAACAAAGAACTCTTTTCAGCAGACATTAGAACTGGATG 329
|||||
QY 3874 ggaagaagattccattatggaggaagagagacccacagcagtcagtcactactctagaa 3933
|||||
Db 330 CCTCAACCTCTGCTGTTCTTGATCA-----TAGCCCTGTGGCTTCAGATATGACACTTA 383
|||||
```



```
Db 1438 TTCCATTTTCCATCTGGTTACTATGACACATCATACATGATTTACTTACCTCAGTA 1497
QY 5062 c---aggcctacccttgcctcaggttcaagaagacagtgagagcgtcattcccgcatctt 5118
Db 1498 CAGAAGAGAATGGCATGGATCAAGAAATCCCCCTTGTGAAGAAGAATCAGCAGTCTCTTG 1557
QY 5119 cttctttgcagaattctcaatatatacaagtggtggtccctcattgggtggtatctcggct 5178
Db 1558 CTTTGTATAAAGACATTCACACAGTATACGGGAAGTGCCTTGAAGAATAACCATCCGGCT 1617
QY 5179 ggtatttgggtcactgaagaatggcatcaccccttatcttcgctgcgtgcattgt 5238
Db 1618 GGCATCTGTGGAGAGTGTCCAGAGTGGATCATCCCTTCCCTGAAGTGTCTGCTTAT 1677
QY 5239 tttccactatttacttgggttaactccgctgaggaactgcataccaattctcagaag 5298
Db 1678 TTTTTCATTACTTAATGGAGTTCCTTCCCCACCCGACATCAAGT--TCCTGGAACAA 1734
QY 5299 gagagtacagtcactctgtagctatctattcttacctacaaatttctcctgctctcc 5358
Db 1735 GCATTTTGAACATTTATGTAGTATCTTCCCTTACCACAAACACCTCATTTGCCCTTTC 1794
QY 5359 aggaatatgggatactgtaaggcccttgcctcagagtggtgagatccctccttac 5418
Db 1795 AAGAAATAGTGAGATAATGAATTAATCACTGATGAAGTGTGGTCCGTAAACAGTGAAGTTA 1854
QY 5419 taaactgtttgaagcaaaaacaccgtggtcaggtacccttagaataaaagtattga 5478
Db 1855 AAAGATATCTAGAAGGTGAAGAGATGCTATAAGATATCCAAAGAGATCTAACAAATAA 1914
QY 5479 tagagttcctgactatagctgcctcctgaatcaagttctctcattcagtgccac 5538
Db 1915 TAAACCTTCAGAGGATTAACAGAGCTCATTAATCAAGCATCCAATTTCTCGTGCCTGA 1974
QY 5539 ggtctcaatgatgacaaagcactcctcctcctcctctctcttcttggtgggtactat 5598
Db 1975 AATCAGGTGTATAGAGCAGAGACCCCAACTCTGTGCTGTGTGCGGATCTCTGCTGT 2034
QY 5599 gttctcagaacattgtgcaggaataattggaacgggagaggttggagcttgctatt 5658
Db 2035 GCTCCAGAGTTACTGCTGCCACTCACTGGAGGGGAGGATGTAGGAGCTGCACAG 2094
QY 5659 ttaacgacttcaactgtgagccgagtgctgtcatttctctaaatacagagaatgccag 5718
Db 2095 CTCACACCTACTCTGTGGCTCTGGAGTGGCATCTTCCGTAGAGTACGGGAATGTCAGG 2154
QY 5719 tggctcgtggtgaaggtaaagccagaggtgtgcctatccagctccttacttggatgaat 5778
Db 2155 TGCTATTTTGTAGTGCAAAACCAAGGCTGTATTTATTTCTCTCCCTTACCTTGATGACT 2214
QY 5779 atggagaacacagacctggcctgaagaggggcaaccccttcttattatctcgtgagcgt 5838
Db 2215 ATGGGAGACCGACACAGGACTCAGAGGGGAATCTTTTACATTTATGCAAGAGCGAT 2274
QY 5839 atcggaagctccatttggctgtgcaacacactgcattatagaagagattgtaggagcc 5898
Db 2275 TCAAGAAGATTCAGAAGCTCTGCGACCAACACAGAGTGTACAGAGGAATTTGGACATGCAC 2334
QY 5899 agagactaatcagatggtatttggattcaactggca 5935
Db 2335 AGGAAGCCAATCAGACACTGGTTGGCATTTGACTGGCA 2371
```

RESULT 13

AC068724/c
LOCUS
DEFINITION Homo sapiens chromosome 15 clone RP11-473C18 map 15q15, complete sequence.
ACCESSION AC068724
VERSION AC068724.7 GI:12739735
KEYWORDS HTG.
SOURCE human.

ORGANISM Homo sapiens
REFERENCE
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 190727)
Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S.,
Dors, M., Dickhoff, R., Fleetwood, P., Harrison, G., Kaur, A., Madan, A.,
Nesbitt, R., Traicoff, R. and Hood, L.
TITLE Sequencing of human chromosome 15 D15S146-D15S117 region
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 190727)
AUTHORS Rowen, L., Madan, A., Qin, S., Abbasi, N., Baradarani, L., Birditt, B.,
Bloom, S., Dors, M., Dickhoff, R., Fleetwood, P., Harrison, G., Kaur, A.,
Madan, A., Nesbitt, R., Traicoff, R., Shaffer, F. and Hood, L.
TITLE Direct Submission
JOURNAL Submitted (08-MAY-2000) Multimegabase Sequencing Center, Institute
for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
98105, USA
3 (bases 1 to 190727)
AUTHORS Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S.,
Dors, M., Dickhoff, R., Fleetwood, P., Harrison, G., Kaur, A., Madan, A.,
Nesbitt, R., Traicoff, R. and Hood, L.
TITLE Direct Submission
JOURNAL Submitted (10-FEB-2001) Multimegabase Sequencing Center, Institute
for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
98105, USA
COMMENT On Feb 10, 2001 this sequence version replaced gi:11138174.
----- Genome Center
Center: Multimegabase Sequencing Center
Center code: UMWSC
Web site: http://chroma.mbt.washington.edu/msg_www
Contact: leerowensystemsbiology.org
----- Summary Statistics
Sequencing vector: pUC18; 108752
Chemistry: Dye-terminator Big Dye; 90% of reads
Chemistry: Dye-primer Big Dye; 10% of reads
Assembly program: Phrap; version 0.990399
Note: Data from overlapping BACs AC009825 [Drafting center: WIBR]
and AC009852 [Drafting center: UMWSC] were added for finishing
----- Location/Qualifiers
1. 190727
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15q15"
/clone="RP11-473C18"
/clone_lib="RPCI human BAC library 11"
/note="This clone overlaps RP11-2016 AC021899, RP11-19D14
AC009825 and RP11-402F9 AC009852. Data from AC009825 and
AC009852 were added and the consensus sequence determined
from RP11-473C18 to the extent possible"
misc_feature 1. 9704
/note="overlap with RP11-2016, AC021899"
42130..42140
/note="low quality data"
unsure 43640..43830
/note="low quality data"
unsure 69905..70000
/note="single read coverage"
misc_feature 87320..190727
/note="overlap with RP11-19D14, AC009825"
unsure 99880..99900
/note="low quality data"
unsure 103370..103380
/note="low quality data"
misc_feature 105748..190727
/note="overlap with RP11-402F9, AC009852"
BASE COUNT 54575 a 40945 c 40610 g 54597 t
ORIGIN
Query Match 5.6%; Score 354; DB 9; Length 190727;

[illegible]

```

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information -----
Center project name: L12436
Center clone name: 473_C_18
-----
* NOTE: This record contains 76 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

```

*	1	be preserved.	716:	contig of 716 bp in length
*				
*	717	816:	gap of	100 bp
*	817	1540:	contig of 724 bp in length	
*	1541	1640:	gap of	100 bp
*	1641	2352:	contig of 712 bp in length	
*	2353	2452:	gap of	100 bp
*	2453	3172:	contig of 720 bp in length	
*	3173	3272:	gap of	100 bp
*	3273	3996:	contig of 724 bp in length	
*	3997	4096:	gap of	100 bp
*	4097	4805:	contig of 709 bp in length	
*	4806	4905:	gap of	100 bp
*	4906	5633:	contig of 728 bp in length	
*	5634	5733:	gap of	100 bp
*	5734	6441:	contig of 708 bp in length	
*	6442	6541:	gap of	100 bp
*	6542	7240:	contig of 699 bp in length	
*	7241	7340:	gap of	100 bp
*	7341	8065:	contig of 725 bp in length	
*	8066	8165:	gap of	100 bp
*	8166	8845:	contig of 680 bp in length	
*	8846	8945:	gap of	100 bp
*	8946	9638:	contig of 693 bp in length	
*	9639	9738:	gap of	100 bp
*	9739	10463:	contig of 725 bp in length	
*	10464	10563:	gap of	100 bp
*	10564	11276:	contig of 713 bp in length	
*	11277	11376:	gap of	100 bp
*	11377	12086:	contig of 710 bp in length	
*	12087	12186:	gap of	100 bp
*	12187	12901:	contig of 715 bp in length	
*	12902	13001:	gap of	100 bp
*	13002	13695:	contig of 694 bp in length	
*	13696	13795:	gap of	100 bp
*	13796	14509:	contig of 714 bp in length	
*	14510	14609:	gap of	100 bp
*	14610	15328:	contig of 719 bp in length	
*	15329	15428:	gap of	100 bp
*	15429	16140:	contig of 712 bp in length	
*	16141	16240:	gap of	100 bp
*	16241	16955:	contig of 715 bp in length	
*	16956	17055:	gap of	100 bp
*	17056	17773:	contig of 718 bp in length	
*	17774	17873:	gap of	100 bp
*	17874	18561:	contig of 688 bp in length	
*	18562	18661:	gap of	100 bp
*	18662	19396:	contig of 735 bp in length	
*	19397	19496:	gap of	100 bp
*	19497	20214:	contig of 718 bp in length	
*	20215	20314:	gap of	100 bp
*	20315	21031:	contig of 717 bp in length	
*	21032	21131:	gap of	100 bp
*	21132	21845:	contig of 714 bp in length	
*	21846	21945:	gap of	100 bp

db_xref="taxon:10090"
/cell_type="embryonal stem"
/dev_stage="embryo"
/chromosome="2"
143..399
/gene="Ubr1"
/number=2
BASE COUNT 273 a 150 c 168 g 319 t
ORIGIN

Query Match 3.4%; Score 215.2; DB 10; Length 910;
Best Local Similarity 89.2%; Pred. No. 3.6e-43;
Matches 232; Conservative 0; Mismatches 28; Indels 0; Gaps 0;
Qy 777 tgggtggatcagcaagtgtgatttttatactgctttcttgcatcatttggcacaattggtg 836
|||||
Db 143 TGGTGGGATCAGCAAGTTGATTCTTCTACTGCTTTCTTACATCATTGGCACAATTAGTG 202
|||||
Qy 837 ccagaaatttacttctgctgaatggaccagacttggaaaagcaggaggaagtgtacaa 896
|||||
Db 203 CCAGAAATTTATTTGCTGAGATGGACCCAGATTTGGAAAAGCAAGAGAGAGTGACAG 262
|||||
Qy 897 atgtcaatattcactccactggaatggatggtacttatttggagaagatccagatatattgctta 956
|||||
Db 263 ATGTCAATACTCACCTCTTGGAGTGGTACTTATTGGAGAGGATCCGGATATTGCTTTA 322
|||||
Qy 957 gagaatttgaagcacagtggagcatttcagcttgtggagggttttcaaaagtggagag 1016
|||||
Db 323 GAGAAATTAACACAGTGGACCGTCCAGTTGTGTGGGAGGTTTTTCAAAAGTGGAGAA 382
|||||
Qy 1017 acaacctattcttcaggga 1036
|||||
Db 383 ACAACATATTCCTGTAGGTA 402

Search completed: March 1, 2002, 16:16:46
Job time: 13031 sec

GenCore version 4.5
Copyright (C) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 1, 2002, 13:51:46 ; Search time 261.48 Seconds
(without alignments)
20682.299 Million cell updates/sec.

Title: US-09-724-126A-1
Perfect score: 6308
Sequence: 1 gcaagaattgcacgagg.....aatttgattgtggtttt 6308

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues
Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
1	4435.4	70.3	6395	20	AAV99308
2	4435.4	70.3	6395	22	AAH14878
3	2546.8	40.4	2550	22	AAH14878
4	996.2	15.8	1001	22	AAH14878
5	996.2	15.8	1001	22	AAH14878
6	813.6	12.9	3327	20	AAH14878
7	813.6	12.9	3327	20	AAH14878
8	656.8	10.4	733	21	AAH14878
9	647	10.3	712	22	AAH14878
10	638.6	10.1	756	21	AAH14878
11	246.8	3.9	807	20	AAH14878

12	241.6	3.8	264	21	AAA00233	Human colon cancer
13	191.6	3.0	800	20	AAV99892	Gastric cancer ass
14	159	2.5	455	20	AAV99674	EST clone DA490.
15	118	1.9	936	22	AAF58252	Oligonucleotide D1
16	118	1.9	936	22	AAF58252	Oligonucleotide D1
17	118	1.9	936	22	AAF58257	Oligonucleotide D1
18	118	1.9	936	22	AAF58259	Oligonucleotide D2
19	118	1.9	936	22	AAF58262	Oligonucleotide D2
20	118	1.9	936	22	AAF58255	Oligonucleotide D1
21	116.6	1.8	308	16	AAAT1876	Human gene signatu
22	116.4	1.8	936	22	AAF58252	Oligonucleotide D1
23	116.4	1.8	936	22	AAF58254	Oligonucleotide D1
24	116.4	1.8	936	22	AAF58257	Oligonucleotide D1
25	116.4	1.8	936	22	AAF58259	Oligonucleotide D2
26	116.4	1.8	936	22	AAF58262	Oligonucleotide D2
27	116.4	1.8	936	22	AAF58255	Oligonucleotide D1
28	105.6	1.7	366	21	AAAC00938	Human secreted pro
29	105.6	1.7	456	21	AAAC00940	Human secreted pro
30	105.6	1.7	1461	21	AAAC78096	Human cancer assoc
31	103.2	1.6	605	21	AAAC00939	Human secreted pro
32	100.8	1.6	602	21	AAAC10733	Human secreted pro
33	100.8	1.6	692	21	AAAC10734	Human secreted pro
34	100.8	1.6	842	21	AAAC10872	Human secreted pro
35	54.8	0.9	2793	21	AAAT0146	Plasmodium falcipa
36	52.4	0.8	244	22	AAF58238	Oligonucleotide D1
37	51.6	0.8	244	22	AAF58238	Oligonucleotide D1
38	49.2	0.8	1727	22	AAH17089	Human cDNA sequenc
39	48.6	0.8	4590	22	AAH24065	Yeast AOP9604-asso
40	46.2	0.7	581	22	AAH09232	Human cDNA clone (
41	45.2	0.7	11722	19	AAV34455	Human MHC class I
42	45	0.7	143068	21	AAF21105	Human low adenosin
43	45	0.7	143068	21	AAF21105	Human low adenosin
44	45	0.7	143068	21	AAAC34983	Human adenosine re
45	45	0.7	143068	21	AAAC35150	Human adenosine re

ALIGNMENTS

RESULT 1
AAV99308
ID AAV99308 standard; cDNA; 6395 BP.
AC AAV99308;
DT 25-MAR-1999 (first entry)
XX cDNA encoding mouse a ubiquitin-protein ligase, Ubql.
DE Ubiquitin-protein ligase; Ubql; mouse; ubiquitinylation; degradation;
KW N-end rule pathway; stress-related muscle wasting; inhibitor; screen; ss.
XX Mus SP.
XX Key Location/Qualifiers
XX CDS 115..5388
XX FT /*tag= a
XX FT /product= Ubql
XX US5861312-A.
XX PN 19-JAN-1999.
XX PD 02-DEC-1997; 97US-0982956.
XX PF 02-DEC-1997; 97US-0982956.
XX PR 02-DEC-1997; 97US-0982956.
XX PA (CALY) CALIFORNIA INST OF TECHNOLOGY.
XX PI Kwon YT, Varshavsky A;
XX WP1; 1999-130395/11.
XX DR P-PSDB; AAW84351.

||||| 3908 ttaataaagaatggggattcaacttttgagttttcaattccatccctgaatttggagttc 3967
QY 4537 agtcttcgattaaattcaaatagcatcaaggaatgggtttattctctttcccaacaa 4596
Db 3968 agtcttcggtgaataattcaaatagtatcaaggaatgggtctctcttcgccacaa 4027
QY 4597 ttatagaattggattgaagtgccactgatgaagggatccctgcagtcgccatgctga 4656
Db 4028 ttacagaattggctgaagtgccctgatgaactgaaccacgagtgcccatgaga 4087
QY 4657 cctggagcactgccttcaactatcaggaactgaataatctattggagatgaagaa 4716
Db 4088 cctggagcagtgctgttcaactccagggcaatcgaaaaactgttggagatgaagaa 4147
QY 4717 aacctctgtttggagcacttcaaataggcagcataatggctgaagacattaatgcagt 4776
Db 4148 aacctctatttggagcacttcaaatagacagcatagcgtctgaaggcgtaatgcagt 4207
QY 4777 ttgcagttgcacagagattacctgtcctcaggtcctgatcacagaacatctgttctc 4836
Db 4208 ttgcagttgcacagaggtacctgtcctcaggtcctgatcacagaacatctgttctc 4267
QY 4837 ttctcagttgttctcctaacaataaatacagaagatcacaccatgcctctgtctatag 4896
Db 4268 tctcgtcagttattcttctaaccctgcaatcagaaaaatacacccaggcctctgtctg 4327
QY 4897 atctgttttcattgtttgggtgctgtgttagcattcccatcctgttatgggatgacc 4956
Db 4328 atcttctccattgtctggctggcgagcttaggttcccatcctgttatgggatgaca 4387
QY 4957 ctgtgtatcgcagcctctctcagttagttcttccataaacacaccttatctctccatt 5016
Db 4388 ccgtggatctgcagcgtgcacacttagttcttcataaacacacctctatcttcccatc 4447
QY 5017 tgatcacatggcaacatgcttcagatatactactatcagtagaca-----cag 5064
Db 4448 tgatcacatggcgacatgcttcagatatactccttacaacagatcacagatctgtctccag 4507
QY 5065 gcttacccttctgcaggttcaagaagacagtggaaggctcattccgcattctcttct 5124
Db 4508 ggcgcgcgttctgtgaggggtgaagagatagtgaggaggtcgtgtgtctcttct 4567
QY 5125 ttgcagaatttctcaatatacagaagtgctccattgggtgtgattatctcgtgctgatt 5184
Db 4568 ttgtggaagtgtgcagcacacacagcgcctcactgggtgcgtgctcccgctgtacc 4627
QY 5185 ttgtgggtctcaactgaagaatggcatcacccttattctcgtgtgctgctattgttttcc 5244
Db 4628 ttgtgctctcctgaggaaaggcatcacccttaccctcctgctgtgctcactgcttttcc 4687
QY 5245 actatttacttgggtgaactccgctgaggaaactgcataccaattctgcagaagagagt 5304
Db 4688 actatttacttgggtgagctccgctgaggaaactgttgcacaattctgctgaaggagaat 4747
QY 5305 acagtgcaactctgtagctatctattacatacaaaattgtctcgtctcttccagggaat 5364
Db 4748 tcagtgcaactctgtagctatctattacatacaaaattgtctcgtctcttccagggaat 4807
QY 5365 attgggtactgtgaaggcccttgcacagaggtggtgtgcagatccctgcttactaaact 5424
Db 4808 attgggtataccataaggcccttactacagaggtggtgtgcagatccctgcttactcaagt 4867
QY 5425 gtttgaagcaaaaaaacacogtgcaggtacccttagaaaaaagaatagttgtatagagc 5484
Db 4868 ctttgaagcagaaaagtgtctgtggtcaggtacccttagaaaaaagaatagttgtatagagc 4927
QY 5485 ttctgatgactatagtgctcctcgtgaatcaaggtctctcatttcaggtgccacaggtctg 5544
Db 4928 ttctgaggactacagctgtcttctaataatcagggtctctcatttaggtgtccacaggtctg 4987
QY 5545 cagatgatgacgaagcatcctgtcctctgcttcttctgtgggctatactatgttctc 5604
|||||

Db 4988 cagatgatgagcaagaagcatcctgtcctctgtctgttttctgtggggccatcctgtgttctc 5047
QY 5605 agaacatttgcycgcaagaaattgtgaacggggaaaggttgagcttgcatttttccag 5664
Db 5048 agaacatttgcycgaagaaattagtgaaatgggaagaggttgagcgtggttttccatg 5107
QY 5665 caattcacttggagccggagctgcatttctcctaagaaatcagagaatgcgagatgggtcc 5724
Db 5108 cgtctcatttgtgtgtgagctgcatttctcctaagaaatcccgagaatgcaggttgggtcc 5167
QY 5725 tggttgaaggttaaaagccagagcgtgtgcctatccagctccttacttggatgaatatggag 5784
Db 5168 tggttgaaggaagaaagccagagcgtgtgcctacccagcccttacttggatgaatatggag 5227
QY 5785 aaacagacccctggcctgaagaggggcaaccccttctatttctgtgcggtatcgga 5844
Db 5228 aaacagacccagggctaaagagaggaaacccacttctatttctcggagcgttatcgga 5287
QY 5845 agtccatttggcttggcaacaacactgcattatagaagagattgctagagcgaagaga 5904
Db 5288 agtccatttggcttggcaacagcactgcattatagaagagattgctcggagccagaga 5347
QY 5905 ctaatcagatgttatttggattcaactggcagttactgtgagctccaactctgcctcaag 5964
Db 5348 ctaatcagatgttatttggatttaactggcagttactgtgagcttcagttcctcctcaag 5407
QY 5965 acaatcacaaatgcagcagtagtaaaagcgtgatttcaaaattatggaatacttctgag 6024
Db 5408 acaatcagtgagatcaataataaagactgactataaaattctagagaacttctgag 5467
QY 6025 gctggaaagtattggaggggttttctgtccatgctcaggttccacttacaataaaaa 6084
Db 5468 acgggggaagtattggaggggttttctgtccatgctcagattccacacacattataaaat 5527
QY 6085 atttcttaagtgaattgtcttcaattagcaaacatagcttccacaggaagaaa-aggac 6143
Db 5528 attctcttaagtaattgtcttcaattatcaaacataagcttcaaggggaaaaacaagac 5587
QY 6144 atagatcaatctgt-----ttt 6160
Db 5588 atagattaatgttttatgttttagaacactaaagaatgcttttccatccaagtgtctat 5647
QY 6161 atgtctagttattccaggaatttattcccttcaataattgtctcatttcttattt 6220
Db 5648 ttctgtaataattccagaaaactccttctcccttcaactgctcctagttccatccat 5707
QY 6221 catccacttggtagatgaagtcacgtccaacagttgtagacattttatgtgtgtttaa 6280
Db 5708 caccacccgttgaatgaggtcacattgaagcatttggacatttctccatctggctaac 5767
QY 6281 tcttctgcaatttctatttgggtttt 6307
Db 5768 atctctgcaccttctatttgggtttt 5794

RESULT 3

AAH14878

ID AAH14878 standard; cDNA; 2550 BP.

XX

AAH14878;

XX

26-JUN-2001 (first entry)

XX

Human cDNA sequence SEQ ID NO:12731.

DE

XX

KW

XX

OS

XX

PN

XX

PD

XX

Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

Homo sapiens.

EP1074617-A2.

07-FEB-2001.


```
QY 2979 gtgacaaagaaggtcacaaatgagagaatcattcacttgcattgcaatgaacccatg 3038
Db 181 gtgacaaagaaggtcacaaatgagagaatcattcacttgcattgcaatgaacccatg 240
QY 3039 ccacacagtgcattgcacaaatctactgagaaatgaaataatgaaactgcttagag 3098
Db 241 ccacacagtgcattgcacaaatctactgagaaatgaaataatgaaactgcttagag 300
QY 3099 aatgtcataaacaagtggccacatttaagaaaccaggtgtatcaggccatggagtttat 3158
Db 301 aatgtcataaacaagtggccacatttaagaaaccaggtgtatcaggccatggagtttat 360
QY 3159 gaactaaagaatgaatcactgaagacttcaatgtactctttatcattactccaaaacc 3218
Db 361 gaactaaagaatgaatcactgaagacttcaatgtactctttatcattactccaaaacc 420
QY 3219 cagcatagaaggctgaacatagcagaagaaaggagaaacaaagaaacaaagatgaa 3278
Db 421 cagcatagaaggctgaacatagcagaagaaaggagaaacaaagaaacaaagatgaa 480
QY 3279 gaattgcgcgcaccaccctcctgaattctgcctgcttttcagcaaaagtgttaacctt 3338
Db 481 gaattgcgcgcaccaccctcctgaattctgcctgcttttcagcaaaagtgttaacctt 540
QY 3339 ctcaactgtatcatgatgtacattctcaggaccgtatttgagcgggcaatagacaca 3398
Db 541 ctcaactgtatcatgatgtacattctcaggaccgtatttgagcgggcaatagacaca 600
QY 3399 gattctaactgtgacgaaggatgtccaaatggcttttcattctgcattgggt 3458
Db 601 gattctaactgtgacgaaggatgtccaaatggcttttcattctgcattgggt 660
QY 3459 ttactagaagaagaacacagcttcaaaaagctcctgaagaagaagtaacatttgacttt 3518
Db 661 ttactagaagaagaacacagcttcaaaaagctcctgaagaagaagtaacatttgacttt 720
QY 3519 tatctaaggcttcaagattgggaagttcagccatgaatatacaaatgcttttgaaaaa 3578
Db 721 tatctaaggcttcaagattgggaagttcagccatgaatatacaaatgcttttgaaaaa 780
QY 3579 ctcaagaagaattcccagttagaagccagagaagcatgataactggagacttcagatg 3638
Db 781 ctcaagaagaattcccagttagaagccagagaagcatgataactggagacttcagatg 840
QY 3639 ttgacacagtgaaagcattgaagaagaaatcttcttaattgtgacacaccatcagga 3698
Db 841 ttgacacagtgaaagcattgaagaagaaatcttcttaattgtgacacaccatcagga 900
QY 3699 tcggaatctattaagaatgatgagattactctatgataaagaagaaagcagaagaaaaa 3758
Db 901 tcggaatctattaagaatgatgagattactctatgataaagaagaaagcagaagaaaaa 960
QY 3759 aaagctgaagctgtaggctacatgcgcagaagatcagc 3799
Db 961 aaagctgaagctgtaggctacatgcgcagaagatcagc 1001
```

```
RESULT 5
AAC86934
ID AAC86934 standard; cDNA; 1001 BP.
XX
AC AAC86934;
XX
XX 02-APR-2001 (first entry)
XX
DE Nucleotide sequence of a partial human Ubrl protein.
XX
KW Ubrl; E3-type protein; ubiquitin system; ubiquitin-protein ligase;
KW N-end rule pathway; intracellular pathogen; Lysteria monocytogenes;
XX Yersinia enterocolitica; muscle wasting; infection; ss.
OS Homo sapiens.
XX
```

```
Key Location/Qualifiers
CDS 1..999
FT /tag= a
FT /product= "Ubrl"
FT /note= "partial sequence"
XX
PN US6159732-A.
XX
PD 12-DEC-2000.
XX
PF 11-JAN-1999; 99US-0228317.
XX
PR 02-DEC-1997; 97US-0982956.
XX
PA (CALY ) CALIFORNIA INST OF TECHNOLOGY.
XX
PI Kwon YT, Varshavsky A;
XX
XX WPI; 2001-090278/10.
DR P-PSDB; AAB31163.
XX
XX Inhibiting the N-end rule pathway in mammalian cells for treating
PT infections and various diseases associated with muscle tissue wasting,
PT by inhibiting the expression of Ubrl gene
XX
PS Claim 4; Column 27-30; 18pp; English.
XX
CC The present sequence encodes a partial Ubrl enzyme. Ubrl is an E3-type
CC protein of the ubiquitin system. Specifically, it is a ubiquitin-protein
CC ligase. The enzyme is specific for destabilising residues exposed at
CC the N-terminus of protein substrates. Inhibition of the expression of
CC Ubrl gene in a cell results in inhibition of the N-end rule pathway.
CC The method is used for treatment of mammalian cells infected with an
CC intracellular pathogen, e.g. Lysteria monocytogenes or Yersinia
CC enterocolitica. Inhibition of N-end rule pathway is also useful for
CC treating various diseases associated with wasting of muscle tissue and
CC infections.
XX
SQ Sequence 1001 BP; 363 A; 186 C; 205 G; 247 T; 0 other;
```

```
Query Match 15.8%; Score 996.2; DB 22; Length 1001;
Best Local Similarity 99.7%; Pred. No. 1.9e-261;
Matches 998; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2799 atggatcccaataagttctgttactgttactcagaggtatgaactgccgaggtttt 2858
Db 1 atggatcccaataagttctgttactgttactcagaggtatgaactgccgaggtttt 60
QY 2859 aacaagaccatattcacaagaaccaggtattgattaaacaataatacataataagaa 2918
Db 61 aacaagaccatattcacaagaaccaggtattgattaaacaataatacataataagaa 120
QY 2919 gaaatgcttcaggtcctcatctatattgtgggtgagcgttatgtacctggagtggaat 2978
Db 121 gaaatgcttcaggtcctcatctattgtgggtgagcgttatgtacctggagtggaat 180
QY 2979 gtgacaaagaaggtgcacaaatgagagaatcattcacttgcattgcaatgaacccatg 3038
Db 181 gtgacaaagaaggtgcacaaatgagagaatcattcacttgcattgcaatgaacccatg 240
QY 3039 ccacacagtgcattgcacaaatctactgagaaatgaaataatgaaactggcttagag 3098
Db 241 ccacacagtgcattgcacaaatctactgagaaatgaaataatgaaactggcttagag 300
QY 3099 aatgtcataaacaagtggccacatttaagaaaccaggtgtatcaggccatggagtttat 3158
Db 301 aatgtcataaacaagtggccacatttaagaaaccaggtgtatcaggccatggagtttat 360
QY 3159 gaactaaagaatgaatcactgaaagacttcaatgtactctttatcattactccaaaacc 3218
Db 361 gaactaaagaatgaatcactgaaagacttcaatgtactctttatcattactccaaaacc 420
```


Db 668. aagatacagactccacactccggtgtgtgctccattctgcctctgttttgcagcctgg 727
QY 3331 ttaacctctcaactgtgatatacatgatgtacattctcagaccgtattttgagcgggcaa 3390
Db 728 ttaacattttgagtcagatgcatgtgtgtcatcatggaacaaattctgcaatggcgctg 787
QY 3391 tagacacagattctcaactgttgagaccgaaggatgctccaaatggcttttcatattcttg 3450
Db 788 tggaaacataatggatgctgctggtcagagtcctatgctgcgaagggtgttacatttaattg 847
QY 3451 cattgggtttactagaagaagcaacagcttcacaaagctcctgaag---aagaagtaa 3507
Db 848 gcatggcactacaagaagaaacacatttagagaatgtccggaagagcatgtagttaa 907
QY 3508 catttgacttttatacagaaggtccaagattgggaagttcagccatgaat-----a 3558
Db 908 catttaccttcaactcagaagatatcaaaacctggtgaagcgccaaaaattctcctagca 967
QY 3559 tacaatgcttttgaaaaaactcaaggaattccccagttagaagccagaggaacatga 3618
Db 968 tactagctatgtggaacacactacaaaaatgctccctacctagaagtcccaagaacatga 1027
QY 3619 taacgtgatacttcagatgtttgacacagtgaaagcgttaagagaaataattctgtttaa 3678
Db 1028 ttcggttgatattgaagacttttaatgctgttaaaagatggaggaga-----gttcac 1081
QY 3679 ttgtagcaaccacatcaggatcggaatctattaaagaatgatgatactcatgataaag 3738
Db 1082 ctaccagtcctgtgcagagacagaaggaaccataatggaagagagttcaaggggcaaaag 1141
QY 3739 aaaaagcagaaagaaagctgaagctgtaggtacatcgccaggaagatcatgg 3798
Db 1142 acaagctgagaggaagaaagacagagattgccagactgcgacagaaagaaatcatgg 1201
QY 3799 ctcaagatgctccttacagaaaaacttcattgaaactcataaaactcatgtatgacaata 3858
Db 1202 ctcatgctgtaaatgcagcgccattttattgatgaacaaagaactctttcagcaga 1261
QY 3859 catcagaatgctgggaagaaagattccattatgaggaagagagcaccacagcagtcac 3918
Db 1262 cattggaactgtagctcaactctgtgtgttcttgatca-----tagcctctgtggctt 1315
QY 3919 gtgaactactagaattgctttgggttcctaaaggggtccatctgttactgaaaaagagg 3978
Db 1316 cagatatgacacttacagcactggcccgcaaaactcaggttctctgaacaaagacaat 1375
QY 3979 tgcctagctgcactcttggccaaagaagaacagagaggtgaaaaatagaaaaataatgcctgg 4038
Db 1376 tcgttacatgtatatgtgtcaagagagcaagaagttaaagtggaaagcagggccaatgg 1435
QY 4039 tattatcgctgtgtccagaataactactgccttaaccacagacaggggaaaaaccatag 4098
Db 1436 tcttggcagcatgtgtcagatgaacaactgtattatcaaaaacagaagaataatttatc 1495
QY 4099 aactctcaggagaagcccttagaccacttttcatggtatccagacttggcatatggaaactt 4158
Db 1496 aagatccagaaaaatat---gatccattattcatgcacccctgatctgtgttgggaacac 1552
QY 4159 atacaggaagctgtgtgtaataatgcacgcaggtgtgctgggagaaagatttttgaagctg 4218
Db 1553 acactagtagctgtgggcacattatgcatgccattgttggcaagggtattttgattccg 1612
QY 4219 tacagctgagctctcagcagcgcatctcatgttgacactttt-----tgacttgg 4266
Db 1613 ttcaagctaaagaacaggaaggaacagagattacgcgttacatcagcgtatgtatgtag 1672
QY 4267 aagtggaagatatcttgccctcttgcgaatctctgtgcataactgtgatcccccatla 4326
Db 1673 aaaaaggaagaaatctcttgcctcttgcgaatctctgtgagatgcttgaactgttattccctgc 1732
QY 4327 ttccctttgcaacctcaagaataaacagtgagaatgcagatgctcttctgtcacttttga 4386

Db 1733 tgcttt-----ctccaagaataatttttaacaacaggtttaaaatttttccagaccacaa 1786
QY 4387 ccttggcactgttgatacagactgttctgtccagaaataatcaggtttaataataagaacatg 4446
Db 1787 atctgactcgtgtgattagacaataatctcagcaataaaaagcattacagtttcttaga 1846
QY 4447 ctaaaagagaaaaacccaattcctatttttcttaatacgaaggaatgggagatttctacttgg 4506
Db 1847 aagaagaaagtactcctaaataatgctctacaaagaattcagaataatgttgatgaattac 1906
QY 4507 agttcattccactcctcgtgattttggcgttgagttctcgtattaaatattcaaatagacaca 4566
Db 1907 agtccctgaagggttcaggcctgattttcgtcctagatcccttattctgagagcataaa 1966
QY 4567 aggaatgtgttattctcttggccacaacaattttatagaattgggattgaaagtgcacactg 4626
Db 1967 aagaataatgcataagcaatttggaaactgtacctacaagggtgggactaaaagttcatocca 2026
QY 4627 atgaaggggactcctcagagtcctcagatgctgaccttgagacacccctgcgttctactatccaag 4686
Db 2027 atgaagaggatcctcgtgttcccaataatgtgtgggtgtagctgcgctacaccatccaaa 2086
QY 4687 caattaaaaactatttggagatgaagaaacccctctgtttggagacacttcaaaatagggc 4746
Db 2087 gcatagaagaatttggagtgaagataaaaccattgtttgttcccttactcttgacagac 2146
QY 4747 agcataatgctcgaagacattaatgagtttcagttgcagtcagacagagatttacctgtccctc 4806
Db 2147 tggatgactgtcttaggttcattgacgagatttgcgcagcacactggacagtggcatcag 2206
QY 4807 aggtcctcgtatacagaanaactctgtgttctctatcagttgttcttcttcttaacataaaat 4866
Db 2207 tttcagtggtgcaggacatttttgaacttttgcatacactgtgtcctaaatgacagcc 2266
QY 4867 cagaagatacacacatgctcctctctctatagatctgtttctcatgttttgggtgtctgtgt 4926
Db 2267 atgagaactcctcagatattagatatgacatgttctatttttgggtgtgtgtgc 2326
QY 4927 tagcattcccatcctgtatttgggatgacctgtgtatctgcagccttcttccagttaatt 4986
Db 2327 ttgcattctcgtgtcagtg-----tcaggatttttccagggtatcagcc 2371
QY 4987 ctctctataaccacttttattctcttccatttgcatacactggcacacatgcttcagatac 5046
Db 2372 ttggcactggagaccttcacatttccattctgttactatgtgcacacatacagatct 2431
QY 5047 tacttaccagtagaacac---aggcctaaccctgtcaggttccaagaagaacagtagaaggg 5103
Db 2432 tacttacctcatgtacagaagaagaatggcatggatcagaanaaatcccccttgtgaagaag 2491
QY 5104 ctcatccgcctctcttcttcttgacagaataattctcacaataacagtggtcccttgggt 5163
Db 2492 aatcagcagttctgtgtgttataaaaacacttccaccagtaacggaaggtgccttgaag 2551
QY 5164 gtgatatcctcgtcgtgtatttgggtctcactgaagaatggcatcaccctctattcttc 5223
Db 2552 aaataccactcggctggcatctgtggagaggtgcagagctgggaatcatgccttctcctga 2611
QY 5224 gctgtgctgcatgttttccactatttacttgggttaactccgcctcgcaggaactgcata 5283
Db 2612 agtgtctgtttatttttcttacttaaatggagtttctctcccaaccgcagactccaag 2671
QY 5284 ccaattctgcagaagagagatagacagtcactgttagctattctattcttaccataaatt 5343
Db 2672 t----tctggaaacaaagccattttgaaacatttatgtagctacttctcccaaccaaaccc 2728
QY 5344 tgttctcgtctctccaggaataatttgggtactgttaagggccttcttccagaggtgtgtgtg 5403
Db 2729 tcaattgccttttccaagaaaaatagtagataatgaattcactgatgtgaagagtggtgccc 2788
QY 5404 cagatcctgcttactaaactgtttgaagcaaaaaaacacccgtgttgcaggttacocataga 5463
Db 2789 gtaacagtgagttaaaagatatctagaagtgaaagagatgctataagatcacaagag 2848

QY 3391 tagacacagattcttaacttgttgaccggaaggatgctccaaatggtcttttctattcttg 3450
Db 843 tggacaataatgatatctgctggtcagagtcocatctgcaagggtgttacatttaattg 902
QY 3451 cattgggtttactagaagaagcaacagcttcacaaagctcctgaag---aagaagtaa 3507
Db 903 gcatggcactacaagaagaaacacatttagagaatgtccggaagagcatgtagtaa 962
QY 3508 catttgaacttttatacaggcttcaagatttgggaagttcagccatgaat-----a 3558
Db 963 catttactctactcagaagatatcaaaacctgtggaagcgccaaaaaattctctagca 1022
QY 3559 tacaatgcttttggaaaaactcaagaatctccocagttagaagcgcoagaaggaatga 3618
Db 1023 tactagctatgtggaaacactacaaaaatgctccctactacagagtcacacaagacatga 1082
QY 3619 taactgataacttcagatgtttgacacagtgaaagcgatgaagataagaaaaattctgttaa 3678
Db 1083 ttcgggtgatatagaagacttttaatgctgttaaaagatgaggaga-----gttcac 1136
QY 3679 ttgtagcaaccacatacaggatcggaatctattaaagaatgagattactcatgataaag 3738
Db 1137 ctaccagctccgtggcagacagacagaagaaccataatggaagagagttcaagggacaaag 1196
QY 3739 aaaaagcagaacgaaaaaagactggaagctgtagctacatcgccgaagaagatcattgg 3798
Db 1197 acaaaagctlgagaggaagaaaaagcagagattgccagactgcccagagaaaaagatcattgg 1256
QY 3799 ctcaagatgtcgttacagaaaaaacttcattgaaactcaataaaactcatgtatgacaata 3858
Db 1257 ctcaagatgtcgaatgcagggcattttattgtgaaacaaagaactcttccagcaga 1316
QY 3859 catcagaatgctcgggaagaaagattccattatgagggaagagagacaccagcagtcga 3918
Db 1317 cattagaactggatgcctcaacctctgtgttcttgatca-----tagcccttgggctt 1370
QY 3919 gtgaactactagaattgctttgggtctctaaacgggtccactctgttactgaaaaaggag 3978
Db 1371 cagatatgacacttacagcaactggcccgccacaaactcaggttctctgaaacaaagacaat 1430
QY 3979 tgcagctgcatctcttgcgaagaagaaacaggaaggtgaaatatagaaaaatagccatgg 4038
Db 1431 tctgtacatgtatatgtgtcaagaggagcaagaagttaaaagtgaagcagggtcaatgg 1490
QY 4039 tattatcgctgtgtccagaatctactgccttaccacagcacaggggaaacaccatag 4098
Db 1491 tcttggcagcatttctcagagatcaactgtattatcaaaaaacagaaagataatttcttc 1550
QY 4099 aactctcaggagaagcccttagaccacattttcatgtgatccagacttggcatatggaactt 4158
Db 1551 aagatccagaaaaatat---gatccattattcatgacacctgactgtctgttggaaacac 1607
QY 4159 atacagggaagctgtgtgtaatgtaagcagagtgctgtggcagaagtagtattttgaagctg 4218
Db 1608 acactagtagctgtgggaacattatgcatgcccattgttggcaagggtattttgattccg 1667
QY 4219 tacagctagctctcagcagcgcattctatgttgacactttt-----tgacttgag 4266
Db 1668 tccaagctaaagaacaggaaggaacacagagattacgcttatacogaagctatgtagtag 1727
QY 4267 aagtgaggaaatatcttgcctctcttgcgaatctctgtgcaatctgtgacactgtgacccattga 4326
Db 1728 aaaaaggagaattctcttgcctcttgcctcttgcctcttgcctcttgcctcttgcctcttgc 1787
QY 4327 ttcctttgcaacctcaaaagataaaacagtgagaatgcagatgctcttgcctcaacttttga 4386
Db 1788 tgcctt-----ctcaagaataatttttaacaacaggttaaaatttttcagaccaaccaa 1841
QY 4387 ccttggcagcgttggtatagacagactgttctggccagaaatatacaggtttataataagaacatg 4446
Db 1842 atctgactcagtggttagaacaatatctcagcaaaataaaagcattacagatttcttagga 1901
QY 4447 ctaaaggagaaaaacccaattctctatttttcttaatacagaagaatggagattctacttgg 4506

Db 1902 aagaagaaagtactctctaaataatgctctctacaaagaattcagaaaaatgttgatgaattac 1961
QY 4507 agtctattccattccctgaggttttgggttgccttgcagcttctctctctctctctctctctct 4566
Db 1962 agtctcctgaaggggttcaaggcctgattttctctctctctctctctctctctctctctct 2021
QY 4567 aggaataatgttattct 4626
Db 2022 aagaataatgtaacagacatttgggaactgctaccctacaagggtgggaactaaaggttcatcca 2081
QY 4627 atgaaggagatcctcagagtcctccatgctgaccttgggagcacctgcgcttttactactacagg 4686
Db 2082 atgaagagatcctcgtgttcccatatgttgggttagctgttgggttagctgcgctacacatccaa 2141
QY 4687 caattgaatactatitgggagatgaagaaaaacctctgttggagcacttcaaaaataggc 4746
Db 2142 gcatagaagaatatttgagtgatgaagataaaacctattgttggctcttactcttgcagac 2201
QY 4747 agcataatggtctgaagcatttaataatgcagtttcagttgcacagagagattacctgtctctc 4806
Db 2202 tggatgactgtcttaggttcattgacgagatttgcgcgacacactggacagtggtcatcag 2261
QY 4807 aggtcctgatacagaagaacatctggttctctctctctctctctctctctctctctctct 4866
Db 2262 tttcagttggtgcaaggacatttttgaaccttttgcacactgttgccttaataatgacagcc 2321
QY 4867 cagaagatacacacatcct 4926
Db 2322 atggaacttccatgcataatgagatattgacatgttctctctctctctctctctctctct 2381
QY 4927 tagcattcccatcctgtatttgggatgacctgttgcctctgcacctcttctcagttagtt 4986
Db 2382 ttgcatttctcgttgcagtg-----tcaggatttttcagggtacagcc 2426
QY 4987 ctctctataaccacctttatct 5046
Db 2427 ttggcactggagaccttccatttccatctgttactatggcacacatcatatacagatct 2486
QY 5047 taottacagtagacac---aggcctacctctgctcaggttccaagaagacagtgaaagg 5103
Db 2487 tactacctcatgtacagaagagaatggcatgatacagaataatccccccttgtgagaag 2546
QY 5104 ctctctccgcatctcttcttcttgcagaataatttctcaatacacaagtggtccctcttgggt 5163
Db 2547 aatcagcagttctgttctgtataaaaaacttccacagtatcacgggaagtgcttgaag 2606
QY 5164 gtgatattcctgctggttatttgggtctctcactgaagaatggcatccacctctctctct 5223
Db 2607 aaataccatccgctggcatctgtggaggaggtgcagagctgggaatcatgctcttctctga 2666
QY 5224 gctgtgctgcatgttttccactatttacttgggttaactccgctcaggaactgcata 5283
Db 2667 agtgtctgctgttatttcttcttacttaaatggagttctctcccccacacatccaa 2726
QY 5284 ccaattctcagaagagagatagctgcactctgtagctatctctctctctctctctctct 5343
Db 2727 t----tctggaaacagccattttgaacattttagtagctatcttctctctctctctct 2783
QY 5344 tgttctgctctctccagggaataatttgggatactgttaagcctcttgcctcagaggtgtgtg 5403
Db 2784 tcaatttgccttttccaagaaaatagtgagataatgaattcactgattgaaaagtggtgccc 2843
QY 5404 cagatcctgcttactaaactgtttgaagcaaaaaacacccgttggcaggttaccctagaa 5463
Db 2844 gtaacagtggaagttaaagatatctagaagtggaagatgctataagatatccaagag 2903
QY 5464 aagaataatgtttgagagcttccctgactatagctgactgctcctcctgaatacaactctc 5523
Db 2904 aatctacaataatlaataaaacctccagagagattacagcagcctcatlaatacaagcatcca 2963
QY 5524 atttcaggtgcccacagcgtctcgcagatgtagcgcaagagcatcctctgctctgctctct 5583

Db 2964 atttctgtgccgaatacagggtgtgataagagcagagcagcccaactctgtgccttggt 3023
 QY 5584 gtggggtatactatgttctcagacaatttctgtccagagaataatgtgaaacggggagagg 5643
 Db 3024 gcgatctctgtgtgtcccgaggttactgtgtccagactgaactggaaggaggatg 3083
 QY 5644 ttggagcttgcatttttcacgcaccttcaactgtgagcggagctgtgeattttctaaaaa 5703
 Db 3084 taggagcctgcacagcgcacaccttactctgtgctctgtgagtggtggcattctctcgagag 3143
 QY 5704 tcagagaatgcggagtggtccctgtgtgaaggtaaagccagaggtgtgcttccatccagctc 5763
 Db 3144 taagggaatgcagtgctatttttagctggaacaaacaaaggctgttttattctctc 3203
 QY 5764 ctacttgatgaatatgggaaacacagaccctggtgagagggggaaccccttcatt 5823
 Db 3204 ctactctgatgactatgggagaccgagggactcagacggggaaatcctttacatt 3263
 QY 5824 tatctgtgagcgtatcggaagctccatttggctgtgagcgaacacactgcattatagaag 5883
 Db 3264 tatgcaagagcgtatcaagagattcagaagctctggcactggaacacacacagtgctcagagg 3323
 QY 5884 agattgtaggagccaagagactaatcagatgttatttggattcaactggca 5935
 Db 3324 aattgacatgcacaggaagccaatcagacactggttggcattgactggca 3375

RESULT 8
 AAA02411
 ID AAA02411 standard; cDNA; 733 BP.
 XX
 AC AAA02411;
 XX
 DT 19-MAY-2000 (first entry)
 XX
 DE Human colon cancer cell line polynucleotide sequence SEQ ID NO:2402.
 XX
 KW Human; colon cancer; tumour; diagnosis; gene expression product;
 KW probe; detection; cancerous state; metastasis; identification;
 KW breast cancer; oestrogen receptor-positive breast cancer; therapy;
 KW oestrogen receptor-negative breast cancer; lung cancer; ss.
 XX
 OS Homo sapiens.
 XX
 XX WO9958675-A2.
 XX
 XX PD 18-NOV-1999.
 XX
 XX PF 13-MAY-1999; 99WO-US10602.
 XX
 XX PR 14-MAY-1998; 98US-0085426.
 XX PR 15-MAY-1998; 98US-0085537.
 XX PR 15-MAY-1998; 98US-0085696.
 XX PR 21-OCT-1998; 98US-0105234.
 XX PR 27-OCT-1998; 98US-0105877.
 XX
 XX (CHIR) CHIRON CORP.
 XX (HYSE-) HYSEQ INC.
 XX
 XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
 XX Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
 XX Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
 XX Leshkowitz D, Kita D, Garcia V, Jones IW, Stache-Crain B;
 XX
 XX WPI; 2000-126369/11.
 XX
 XX Polynucleotide library used to determine cancerous states of mammalian
 XX cells -
 XX
 XX Claim 1; Page 958; 1097pp; English.
 XX
 XX AAA00010 to AAA02716 represent polynucleotides isolated from cDNA
 CC libraries constructed from human colon cancer cell lines. The present

CC invention also describes a method of detecting differentially expressed
 CC genes correlated with a cancerous state of a mammalian cell, comprising
 CC detecting at least one differentially expressed gene product in a test
 CC sample derived from a cell suspected of being cancerous, where detection
 CC of the differentially expressed gene product is correlated with a
 CC cancerous state of the cell from which the test sample was derived.
 CC The polynucleotides sequences can be used in a method for detecting
 CC differentially expressed genes correlated with a cancerous state of a
 CC mammalian cell. The polynucleotides can also be used as probes for
 CC detecting and mapping related genes. They can be used in diagnosis and
 CC prognosis of diseases and disorders (e.g. identification of
 CC pre-metastatic or metastatic cancerous states, stages of cancer, or
 CC responsiveness of cancer to therapy). This is particularly for breast
 CC cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
 CC negative breast cancer, lung cancer, and colon cancer.
 XX
 SQ Sequence 733 BP; 218 A; 145 C; 152 G; 210 T; 8 other;

Query Match 10.4%; Score 656.8; DB 21; Length 733;
 Best Local Similarity 97.5%; Pred. No. 7e-169;
 Matches 696; Conservative 0; Mismatches 13; Indels 5; Gaps 3;

QY 1405 catatgaccacgtcattacacagctcaacagagctcttgaactgtgagctgcagagccc 1464
 Db 20 cagcagaccacgtcattacacagctcaacagagctcttgaactgtgagctgcagagccc 79
 QY 1465 agttgcataccactgcattgacaaagagggtcgtcggtcttaagcgaggacttatg 1524
 Db 80 agttgcnttccactgcatgacaaagagggtcgtcggtcttaagcgaggacttatg 139
 QY 1525 ctgcttgcaggagcaaggaagatatataagagtcattcagaaaaatgtctctcaacatc 1584
 Db 140 ctgcttgcaggagcaaggaagatttaaaagagtcattcagaaaaatgtctctcaacatc 199
 QY 1585 caactcatgataagattacacacagagatttggtcctcagaaaatttgcttggctc 1644
 Db 200 caactcatgataagattacacacagagatttggtcctcagaaaatttgcttggctc 259
 QY 1645 ttggtctctggatgaacaaatttatgactattcaagtcagtcagtcagtcagtcagtcagtc 1704
 Db 260 ttgg-tcttgatgaacaaatttatgactattcaagtcagtcagtcagtcagtcagtcagtc 318
 QY 1705 aagcatgcttagagaagaacactgactcggagaaatccctgtctcctaagcagggttaatgc 1764
 Db 319 aagcatgcttagagaagaacactgactcggagaaatccctgtctcctaagcagggttaatgc 378
 QY 1765 ttgggagtcacaaagctttataaagggtgcgcgtaagatcccttcattgatgtattcagca 1824
 Db 379 ttgggagtcacaaagctttataaagggtgcgcgtaagatcccttcattgatgtattcagca 438
 QY 1825 gttttttatgagatggaatacaaaaaactctttgtctatggaattgtgaagtattata 1884
 Db 439 gttttttatgagatggaatacaaaaaactctttgtctatggaattgtgaagtattata 498
 QY 1885 aacaactgcagaaagaataatatacagtgatgatcatgacagaagtcattctataactgcac 1944
 Db 499 aacaactgcagaaagaataatatacagtgatgatcatgacagaagtcattctataactgcac 558
 QY 1945 ttacagttcagatgtttactgttctactctgtgctgcagacattctattgaagagcagaatg 2004
 Db 559 ttacagttcagatgtttactgttctactctgtgctgcagacattctattgaagagcagaatg 618
 QY 2005 ttatctctgtcattactgaacactctgctagaagttttacotcgtactgttggacagg--aa 2062
 Db 619 ttatctctgtcattactgaacactctgctagaagttttacotcgtactgttggacagg--aa 678
 QY 2063 caataaaatccacttccagggttatagcc--aggacaaaattgggaagagatat 2114
 Db 679 caataaaatccacttccagggttatagcc--aggacaaaattgggaagagatat 732

RESULT 9

	AAH07621	ID.	AAH07621 standard; cDNA; 712 BP.	
	XX AC	AAH07621;		
QY	XX	26-JUN-2001 (first entry)		
DT	XX	Human cDNA clone (5'-primer) SEQ ID NO:4456.		
DE	XX	Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.		
KW	XX	Homo sapiens.		
OS	XX	EP1074617-A2.		
PN	XX	07-FEB-2001.		
PD	XX	28-JUL-2000; 2000EP-0116126.		
PF	XX	29-JUL-1999; 99JP-0248036.		
PR	XX	27-AUG-1999; 99JP-0300253.		
PR	XX	11-JAN-2000; 2000JP-0118776.		
PR	XX	02-MAY-2000; 2000JP-0183767.		
PR	XX	09-JUN-2000; 2000JP-0241899.		
PR	XX	(HELI-) HELIX RES INST.		
PA	XX	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;		
PI	XX	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;		
PT	XX	WPI; 2001-318749/34.		
DR	XX	Primer sets for synthesizing polynucleotides, particularly the 5602		
PT	XX	full-length cDNAs defined in the specification, and for the detection		
PT	XX	and/or diagnosis of the abnormality of the proteins encoded by the		
PT	XX	full-length cDNAs -		
PS	XX	Claim 1; SEQ ID 4456; 2537pp + CD ROM; English.		
CC	XX	The present invention describes primer sets for synthesising 5602		
CC	XX	full-length cDNAs defined in the specification. Where a primer set		
CC	XX	comprises: (a) an oligo-dr primer and an oligonucleotide complementary		
CC	XX	to the complementary strand of a polynucleotide which comprises one of		
CC	XX	the 5602 nucleotide sequences defined in the specification, where the		
CC	XX	oligonucleotide comprises at least 15 nucleotides; or (b) a combination		
CC	XX	of an oligonucleotide comprising a sequence complementary to the		
CC	XX	complementary strand of a polynucleotide which comprises a 5'-end		
CC	XX	sequence and an oligonucleotide comprising a sequence complementary to a		
CC	XX	polynucleotide which comprises a 3'-end sequence, where the		
CC	XX	oligonucleotide comprises at least 15 nucleotides and the combination of		
CC	XX	the 5'-end sequence/3'-end sequence is selected from those defined in		
CC	XX	the specification. The primer sets can be used in antisense therapy and		
CC	XX	in gene therapy. The primers are useful for synthesising polynucleotides,		
CC	XX	particularly full-length cDNAs. The primers are also useful for the		
CC	XX	detection and/or diagnosis of the abnormality of the proteins encoded by		
CC	XX	the full-length cDNAs. The primers allow obtaining of the full-length		
CC	XX	cDNAs easily without any specialised methods. AAH03166 to AAH13628 and		
CC	XX	AAH13633 to AAH18742 represent human cDNA sequences; AB92446 to		
CC	XX	AB95893 represent human amino acid sequences; and AAH13629 to AAH13632		
CC	XX	represent oligonucleotides, all of which are used in the exemplification		
CC	XX	of the present invention.		
XX	XX	Sequence 712 BP; 229 A; 132 C; 156 G; 190 T; 5 other;		
Query Match		10.3%; Score 647; DB 22; Length 712;		
Best Local Similarity		97.6%; Pred. No. 3.3e-166;		
Matches		687; Conservative 0; Mismatches 13; Indels 4; Gaps 3;		
QY	YY	1195aacctggaagacagggtactataaaagaattcacgcgtgccgttgatgaagggtaa	1254	
DG	YY			
DG	YY	8acctgggaagcagggtactataaaagaattcacgcgtgccgttgatgaagggtaa	67	

XX (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
XX
XX WPI; 2000-126369/11.
DR
XX
XX Polynucleotide library used to determine cancerous states of mammalian
PT cells -
PT
XX
XX Claim 1; Page 916-917; 1097pp; English.
PS
XX
XX AAA00010 to AAA02716 represent polynucleotides isolated from cDNA
CC libraries constructed from human colon cancer cell lines. The present
CC invention also describes a method of detecting differentially expressed
CC genes correlated with a cancerous state of a mammalian cell, comprising
CC detecting at least one differentially expressed gene product in a test
CC sample derived from a cell suspected of being cancerous, where detection
CC of the differentially expressed gene product is correlated with a
CC cancerous state of the cell from which the test sample was derived.
CC The polynucleotides sequences can be used in a method for detecting
CC differentially expressed genes correlated with a cancerous state of a
CC mammalian cell. The polynucleotides can also be used as probes for
CC detecting and mapping related genes. They can be used in diagnosis and
CC prognosis of diseases and disorders (e.g. identification of
CC pre-metastatic or metastatic cancerous states, stages of cancer, or
CC responsiveness of cancer to therapy). This is particularly for breast
CC cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
CC negative breast cancer, lung cancer, and colon cancer.
XX
XX Sequence 756 BP; 219 A; 145 C; 154 G; 219 T; 19 other;

Query Match 10.1%; Score 638.6; DB 21; Length 756;
Best Local Similarity 95.9%; Pred. No. 6.7e-164;
Matches 677; Conservative 0; Mismatches 26; Indels 3; Gaps 3;

QY 1405 catatgacacgctatatacagctcaacagagctcttgactgtgactgcagagggccc 1464
DB || |||||
DB 54 cagcagaccacgctatatacagctcaacagagctcttgactgtgactgcagagggccc 113
QY 1465 agttgcataccctgccattgacaaagaggggtcgtcggtcttaaagcggagcttatg 1524
DB |||||
DB 114 agttgcataccctgccattgacaaagaggggtcgtcggtcttaaagcggagcttatg 173
QY 1525 ctgcttgccaggagcaaaagagatataaagagctcattcagaaaatgtctcacaacatc 1584
DB |||||
DB 174 ctgcttgccaggagcaaaagagatataaagagctcattcagaaaatgtctcacaacatc 233
QY 1585 caactcatgtagaagtattacatcagagattatggctcatcagaaaattgtcttgctgc 1644
DB |||||
DB 234 caactcatgtagaagtattacatcagagattatggctcatcagaaaattgtcttgctgc 293
QY 1645 ttggttcctgtagaacaataattatgaacttcaagctcattgagcagatctttgcc 1704
DB |||||
DB 294 tnggttcctgtagaacaataattatgaacttcaagctcattgagcagatctttgcc 353
QY 1705 aagcatgcttagaagaacacctgactcggagaaatccctgtctcataagcaggttaatgc 1764
DB |||||
DB 354 aagcatgcttagaagaacacctgactcggagaaatccctgtctcataagcaggttaatgc 413
QY 1765 ttgggatgcaagcctttataaaggtgcccgtaagatcccttcattgaattgtcttcagca 1824
DB |||||
DB 414 ttgggatgcaagcctttataaaggtgcccgtaagatcccttcattgaattgtcttcagca 473
QY 1825 gtttttttggagatggaatacaaaaactctttgtgtatggaattgtgaagtattata 1884
DB |||||
DB 474 gtttttttggagatggaatacaaaaactctttgtgtatggaattgtgaagtattata 533

QY 1885 acaactgcagaaagaatatatactcaagtgcagatgcacagagaagtatactataactgcac 1944
DB |||||
DB 534 acaactgcagaaagaatatatactcaagtgcagatgcacagagaagtatactataactgcac 593
QY 1945 ttctcagttcagatgtttactgttctcactctgctgcacatcttattgaaagcagaatg 2004
DB |||||
DB 594 ttctcagttcagatgtttactgttctcactctgctgcacatcttattgaa-aacagaatg 651
QY 2005 ttctcagttcagatgtttactgttctcactctgctgcacatcttattgaaagcagaatg 2064
DB |||||
DB 652 ttctcagttcagatgtttactgttctcactctgctgcacatcttattgaaagcagaatg 711
QY 2065 ataaattcaactcaggtttatagccagacaaaattgggaagatg 2110
DB |||||
DB 712 ntaaatnacttccagggttat-gcnggacacattggnagatt 756
RESULT 11
AA339891
ID AA339891 standard; DNA; 807 BP.
XX
AC AA339891;
XX
DT 02-JUL-1999 (first entry)
XX
XX Gastric cancer associated gene.
XX
XX Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.
XX
OS Homo sapiens.
XX
PN WO9904265-A2.
XX
PD 28-JAN-1999.
XX
PF 15-JUL-1998; 98WO-US14679.
XX
PR 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX
DR WPI; 1999-132448/11.
XX
XX New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
PS
XX Claim 67; Page 586-587; 787pp; English.
XX
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and

Db	639	TTAAAGATTCTTAGAAGGTGAAACAGAGATGCTATTAAGATATCCACAGAGAAATCTTAACAA	580
Qy	5474	tttgatagagcttctctgactagctactagctcctccctgaaatcaagcttctctcatttcagggtg	5533
Db	579	ATTATAAACCCTCCAGAGGATTACAGACGCCTCAATTAATCAAGCATCCAATTTCTCGTG	520
Qy	5534	cccacggtctcgaagtatgagcgaaagcatcctctgctctctgcttctctgtggtggtat	5593
Db	519	CCCGAAATCAGGTGGTGATTAAGACAGACAGCCCAACTCTGTGCTTGTGTGCGGATCTCT	460
Qy	5594	actatgtctcagaaacatttctgccaggaataattgtgaacggggaagaggttgagacttg	5653
Db	459	GCTGTGCTCCACAGATTACTGCTGCCAGACTGAACCTGGAAGGGGAGGATGTAGGAGCCTG	400
Qy	5654	catttttcacgcacttcaactgtggagccggagctctgcatttttcttaaaatcagagaatg	5713
Db	399	CACAGCTCACACCTACTCCTGTGGCTCTGGAGTGGGCATCTTCTGAGAGTACGGGAATG	340
Qy	5714	ccgagtggtcctcgttgtaaggtaaaagccagagcgtgtgacctatccagctctccttacttggg	5773
Db	339	TCAGTGTCTATTTTAGCTGGCAAAACCAAGGCTGTTTTTATTCTCTCTCTTACCTTGA	280
Qy	5774	tgaatatagaataaacagaccctcgtcctgaagaggggcaaccccttcatttctctgtga	5833
Db	279	TGACTATGGGAGACCGACACGGGACTCAGACGGGAAATCCTTTACATTTATGCAAGA	220
Qy	5834	gcggtatcggagctccatttggctgtcgtggaacaacacactgcattatagaagagattgctag	5893
Db	219	CGGATTCAAGAAGATTACAGAAGCTCTGGCACCAACACACAGTGTACAGAGGAAATTGGACA	160
Qy	5894	gagccaagagactaatcagatgttatttggattccaactggca	5935
Db	159	TGCACAGGAAGCAATCAGACACTGGTTGGCATTGACTGGCA	118
RESULT 14			
AAV89674/c			
ID	AAV89674 standard; cDNA; 455 BP.		
XX	AAV89674;		
XX	15-FEB-1999 (first entry)		
DT	EST clone DA490.		
DE	Human; secreted protein; expressed sequence tag; EST; haematopoiesis;		
KW	tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic		
KW	receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumor		
KW	gene therapy; ss.		
XX	Homo sapiens.		
OS	W09845436-A2.		
PN	15-OCT-1998.		
PD	10-APR-1998; 98WO-US06955.		
PF	10-APR-1997; 97US-0838821.		
PR	(GEMV) GENETICS INST INC.		
XX	Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;		
XX	Racie LA, Spaulding V, Treacy M;		
PI	WPI; 1999-070077/06.		
XX	New polynucleotides encoding human secreted proteins - derived from		
PT	e.g. human blood, kidney, foetal lung, placenta, testes, brain,		
PT	ovary, pituitary, retina and colon cDNA libraries.		
XX	Claim 1; Page 293; 618pp; English.		
PS			
XX			

CC	The present sequence represents a human expressed sequence tag (EST).
CC	The polynucleotide, which is a secreted EST, and the encoded protein
CC	are predicted to have useful biological activities which would make
CC	them suitable for treating, preventing or ameliorating medical
CC	conditions in humans and animals, although no supporting data is
CC	given. Suggested activities include nutritional activity, immune
CC	stimulating or suppressing activity, haematopoiesis regulating
CC	activity, tissue growth activity, activin/inhibin activity,
CC	chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC	activity, receptor/ligand activity, anti-inflammatory activity,
CC	cadherin/tumour invasion suppressor activity, tumour inhibition
CC	activity. The polynucleotide may also be useful for gene therapy.
XX	
SQ	Sequence 455 BP; 108 A; 121 C; 107 G; 119 T; 0 other;
	Query Match 2.5%; Score 159; DB 20; Length 455;
	Best Local Similarity 61.9%; Pred. No. 4.3e-33;
	Matches 252; Conservative 0; Mismatches 155; Indels 0; Gaps 0;
QY	5529 aggtgcacaggtctgcagatgagcagaaagcatctgcctctcctcttctctgtgg 5588
Db	451 AGGTGCCGAATTCAGGTGGTGATAAGAGCAGACCCCACTCTGTGCTTGTGCGGA 392
QY	5589 gctatactatgtctcagaacatttgtccagagaaatttgaaacggggaagaggttga 5648
Db	391 TCTGTCTGTGCTCCACAGAGTTACTGTGCCAGACTGAACCTGGAAGGGAGGATGTAGGA 332
QY	5649 gcttgcaattttacgcacctcaactcgttgagccggagctgcatttctctaaaaatcaga 5708
Db	331 GCCTGSCAGAGTCAACACTACTCTGTGGCTCTGGAGTGGGCATCTTCTCGAGAGTACGG 272
QY	5709 gaatgcgcagtggtccctgggtgaaggtaaagccagagagctgtgcctatccagctccctac 5768
Db	271 GAATGTCAGGTGCTATTTTAGCTGGCAAAACCAAGCGCTGTTTATCTCCCTCCCTTAC 212
QY	5769 ttgatgaatatgagaaacagacccttggcctgaagagggggcaaccccttcattatct 5828
Db	211 CTTGATGATTATGGGAGACCGACAGGAGCTCAGACGGGGAATCCTTTACATTATGC 152
QY	5829 cgtgagcggatcgcgaagctccatttggctggaacaaacactgcattatagaagagatt 5888
Db	151 AAAGAGCGGATTCAAGAAGATTCAAGAGCTCTGGCAACCAACACAGTGTGCACAGAGGAATT 92
QY	5889 gctaggaccaagagactaatcagatgtatttggattcaactggca 5935
Db	91 GGACATGCACAGGAAGCAATCAGACACTGTTGGCATTTGACTGGCA 45
	RESULT 15
	AAF58252/c
ID	AAF58252 standard; DNA; 936 BP.
XX	
AC	AAF58252;
AC	
DT	24-APR-2001 (first entry)
XX	
XX	Oligonucleotide D1835.
DE	
XX	
KW	Electron-transfer group; ETM; mismatch; genotyping;
KW	gene expression; ss.
XX	
OS	Synthetic.
XX	
PN	WO200107665-A2.
XX	
PD	01-FEB-2001.
XX	
PF	26-JUL-2000; 2000WO-US20476.
XX	
PR	26-JUL-1999; 99US-0145695.
PR	17-MAR-2000; 2000US-0190259.
XX	

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 1, 2002, 12:40:50 ; Search time 95.27 Seconds
(without alignments)
14995.520 Million cell updates/sec

Title: US-09-724-126A-1
Perfect score: 6308
Sequence: 1 gccagaattcgccagcagg.....aatttgatttggtgtttt 6308

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4435.4	70.3	6395	2 US-08-982-956-1	Sequence 1, Appli
2	4435.4	70.3	6395	3 US-09-228-317-1	Sequence 1, Appli
3	996.2	15.8	1001	2 US-08-982-956-2	Sequence 2, Appli
4	996.2	15.8	1001	3 US-09-228-317-2	Sequence 2, Appli
5	57.4	0.9	7218	1 US-08-232-463-14	Sequence 14, Appl
6	52.2	0.8	7218	1 US-08-232-463-14	Sequence 14, Appl
7	44.4	0.7	7333	3 US-09-138-024-21	Sequence 21, Appl
8	44.2	0.7	72928	3 US-09-009-913-1	Sequence 1, Appli
9	42.2	0.7	325	4 US-08-991-789A-236	Sequence 236, App
10	41	0.6	1517	3 US-08-963-602-6	Sequence 6, Appli
11	40.4	0.6	152331	3 US-09-128-155-16	Sequence 16, Appl
12	39.8	0.6	2861	1 US-08-299-953-1	Sequence 1, Appli
13	39.8	0.6	2861	1 US-08-459-415-1	Sequence 1, Appli
14	39.8	0.6	2861	5 PCT-US95-11231-1	Sequence 1, Appli
15	39.8	0.6	3881	1 US-08-299-953-2	Sequence 2, Appli
16	39.8	0.6	3881	1 US-08-459-415-2	Sequence 2, Appli
17	39.8	0.6	3881	5 PCT-US95-11231-2	Sequence 2, Appli
18	39.8	0.6	4376	1 US-08-119-125A-1	Sequence 1, Appli
19	38.8	0.6	246240	2 US-08-724-394A-20	Sequence 20, Appl
20	38.8	0.6	246240	2 US-08-724-394A-21	Sequence 21, Appl
21	38.8	0.6	246240	2 US-08-724-394A-22	Sequence 22, Appl
22	38.4	0.6	5703	4 US-09-280-590A-36	Sequence 36, Appl
23	38.4	0.6	18596	4 US-09-318-448-11	Sequence 11, Appl
24	37.8	0.6	19011	1 US-08-310-356-36	Sequence 36, Appl
25	37.8	0.6	19557	5 PCT-US92-06300-1	Sequence 1, Appli
26	37.6	0.6	72928	3 US-09-009-913-1	Sequence 1, Appli
27	37.4	0.6	1333	4 US-08-543-246B-15	Sequence 15, Appl

c 28	37.4	0.6	1387	4 US-08-543-246B-1	Sequence 1, Appli
c 29	37	0.6	1113	1 US-08-341-538A-1	Sequence 1, Appli
c 30	37	0.6	1113	2 US-08-725-518-1	Sequence 1, Appli
c 31	37	0.6	1618	3 US-08-889-108-1	Sequence 1, Appli
c 32	37	0.6	1618	3 US-08-889-108-3	Sequence 3, Appli
c 33	37	0.6	1618	4 US-08-120-601B-1	Sequence 1, Appli
c 34	37	0.6	1618	4 US-08-120-601B-3	Sequence 3, Appli
c 35	37	0.6	1618	5 PCT-US94-10358-1	Sequence 1, Appli
c 36	37	0.6	1618	5 PCT-US94-10358-3	Sequence 3, Appli
c 37	37	0.6	4875	1 US-08-460-739-1	Sequence 1, Appli
c 38	36.6	0.6	1480	4 US-09-290-640-65	Sequence 65, Appl
c 39	36.6	0.6	3784	1 US-07-623-033-1	Sequence 1, Appli
c 40	36.4	0.6	1222	4 US-08-543-246B-5	Sequence 5, Appli
c 41	36.2	0.6	9636	1 US-08-323-170B-1	Sequence 1, Appli
c 42	36.2	0.6	19124	2 US-08-487-826B-13	Sequence 13, Appl
c 43	35.4	0.6	252	2 US-08-332-766A-28	Sequence 28, Appl
c 44	35.4	0.6	2135	4 US-08-430-286A-1	Sequence 1, Appli
c 45	35.2	0.6	289	4 US-09-007-005-17	Sequence 17, Appl

ALIGNMENTS

RESULT 1
US-08-982-956-1
; Sequence 1, Application US/08982956
; Patent No. 5861312
; GENERAL INFORMATION:
; APPLICANT: Varshavsky, Alexander
; APPLICANT: Kwon, Yong Tae
; TITLE OF INVENTION: NUCLEIC ACID ENCODING MAMMALIAN UBRI
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kevin M. Farrell, P. C.
; STREET: P.O. Box 999
; CITY: York Harbor
; STATE: ME
; COUNTRY: US
; ZIP: 03911
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/982.956
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Farrell, Kevin M.
; REGISTRATION NUMBER: 35,505
; REFERENCE/DOCKET NUMBER: CIT-2001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (207) 363-0558
; TELEFAX: (207) 363-0528
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6395 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 115..5385
; US-08-982-956-1

Query Match 70.3%; Score 4435.4; DB 2; Length 6395;
Best Local Similarity 86.3%; Pred. No. 0;
Matches 4997; Conservative 0; Mismatches 721; Indels 69; Gaps 6;
Oy 590 ttcaggggccgtctgtaaaagtgtctgtctctc-tccgacggccacaggtttccgct 648

Db 4388 CCGTGGATCTGCAGCGCTCGCACCTTAGTCTTCATATAAACACCTCTATCTCTCCATC 4447
Qy 5017 tgatcacatggcacacatgcttcagatactactactactacagtagagaca-----cag 5064
Db 4448 TGATCACCATTGGCGCACATGCTTTCAGATACCTCTTACACACAGATACAGATCTGTCTCCAG 4507
Qy 5065 gcctacccttgctcaggttcaagaagacagtgagagcctcaatccgcacatcttctct 5124
Db 4508 GCCGCCCTGCTGCTGAGGGTGAAGAGGATAGTGAAGAGCTCGCTGTGCATCTGCTTCT 4567
Qy 5125 ttgcagaattctcaatacaaatggctccatgggtgtgatatctctgctgctgatt 5184
Db 4568 TTGTGGAAGTGTGCGACGACACAGACGCGCTCACCTGGGTGCGGTCTCCGCGCTGGTACC 4627
Qy 5185 tgtgggttactgaagaatggcaccaccccccttatctctgctgtgctgattgttttcc 5244
Db 4628 TGTGCTCTCCCTGAGGAAGGCATCACCCCTTACCTCCGCTGTGCTGCACTGCTTTTCC 4687
Qy 5245 actatttacttgggttaactccgctgaggaactgcataccaattctgcagaaggagat 5304
Db 4688 ACTATTACTTGGAGTAGTCCGCTGAAGAACTGTTTGCCAAATCTGCTGAAGGAGAAT 4747
Qy 5305 acagtgcactctgagctatctattacaccataaattgttctcctgctctccaggaat 5364
Db 4748 TCAGTGCACCTGTGACGTATCTTATCCACCAAAATTTGTTCTGCTTTTCCAGGAAT 4807
Qy 5365 attgggtactgaaggcccttgcctccagagtggtgtagacatcctgcttactactaaat 5424
Db 4808 ATTGGGATACCAATAAGGCCCTTACTACAGAGGTGCTGTGGAGACTCTGCTTACTCAAGT 4867
Qy 5425 gtttgaagcaaaaaaacccgtggtcaggtacccctagaaaaaagaataagttgatagac 5484
Db 4868 CTTTGAAGCAGAAAAAGTGTGTGGTACAGGTACCTAGAAAAAGAAATAGTTGATAGAC 4927
Qy 5485 ttcctgtagactatagctgctccctgaatcaagcttctcattcaagggtgccacaggtcg 5544
Db 4928 TTCTTGAGGACTACAGCTGTCTTAAATCAGGCTTCTCATTAGGTGCCACGGCTCG 4987
Qy 5545 cagatgtagcgaagcactcctcctgctccttcttctgtgggctatactatgttctc 5604
Db 4988 CAGATGATGAGCGAAAGCATCTGCTCTGTCTTCTGTGGGCCATCCTGTGTCTC 5047
Qy 5605 agaacattgctgcaggaataattgtaacggggaagaggttgagagcttgctattttcag 5664
Db 5048 AGAACATCTGTGCGCAAGAAATAGTGAATGGGAAGAGGTTGGAGCGTTCGTTTCATG 5107
Qy 5665 cacttcaactgtgagccgagctgcattttctcctaaaaatacagagaatccgagtggtcc 5724
Db 5108 CGCTTCATTGTGTGCTGGAGTCTGCATTTCTTAAATCCGAGAAATCAGGGTGGTCC 5167
Qy 5725 tgggtgaaggttaagcagagcgtgtgcctatccagctccttacttggtgatgaatggag 5784
Db 5168 TGGTGAAGGAAAGCCAGAGGCTGTGCTTACCAGCCCTTACTTGGATGAATATGGAG 5227
Qy 5785 aaacagacctgacctgaagaggggcaacccctcatttctctgtagcggtatcgga 5844
Db 5228 AAACAGACCCAGGCTAAAGAGAGGAAACCCACTTTCATTATCTCGGAGCGGTATCGGA 5287
Qy 5845 agctccatttggctggaacacacatgcattatagaagagattgctaggagccaagaga 5904
Db 5288 AGCTGCATTTGCTGCGCAACAGCACTGCATTATAGAAGAGATTGCTCGGAGCCAGGAGA 5347
Qy 5905 ctaatcagatgtatttggattcaactggcagttactgagctccaactctgcctcaag 5964
Db 5348 CTAATCAGATGCTATTGTGATTGATGCTGAGTTACTCTGAGCTTCAAGTTCGCCCTCAAG 5407
Qy 5965 acaatcacaaatgacgacagtagtaaaaggctgattcaaaattatggaactttctgagg 6024
Db 5408 ACAATCATGAGTGACATCAATAAAGACTGATCTAAATTTCTAGAGAACCTTCTGAGG 5467
Qy 6025 gctgggaagtagttgaggggtcttttgcctcatgtgccaggttcaacttacatcaataaat 6084

Db 5468 ACGGGGGAAGTATTGGAGGGTCTTTTGTATCCATGTCCAGATTCCACACACATTAATAAAAT 5527
Qy 6085 atttcttaatgaggtatttcttccaattagcaaaaacataatgcttcaacagggaaaaa-aggac 6143
Db 5528 ATTCCCTTAATGGAATATTGCTTTTCAATTAATAAACAATTAAGCTTCAAGGGAACAAAGAC 5587
Qy 6144 atagatcaaatctgt-----ttt 6160
Db 5588 ATAGATTAATGTTTTATGTCTTAGAACACTAAAGAAATGCTTGTTCATCCAAAGTGTCTAT 5647
Qy 6161 atgtgtagtatttccaggaatttattccctctcaataatttgcctcatttcttatttttt 6220
Db 5648 TTCTGCTAATATTTCCAGAAACTCTTTTCCCTTCATAACTGCTCTAGTTTCATTTCATAT 5707
Qy 6221 catccacttggtagtagaagtcacgtcaaacagcttgtagacattttatgcttgggttaac 6280
Db 5708 CACCCACCTGGTTAATGAGGTACATTAAGCATTTTGTGGACATTTCTCCATCTGGCTAAC 5767
Qy 6281 tcttctgcaatttctgatttgggttt 6307
Db 5768 ATCTCTGCACCTTTGTATTGGTGT 5794

RESULT 2
US-09-228-317-1
; Sequence 1, Application US/09228317
; Patent No. 6159732
; GENERAL INFORMATION:
; APPLICANT: Varshavsky, Alexander
; APPLICANT: Kwon, Yong Tee
; TITLE OF INVENTION: NUCLEIC ACID ENCODING MAMMALIAN UBRL
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kevin M. Farrell, P.C.
; STREET: P.O. Box 999
; CITY: York Harbor
; STATE: ME
; COUNTRY: US
; ZIP: 03911
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/228,317
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Farrell, Kevin M.
; REGISTRATION NUMBER: 35,505
; REFERENCE/DOCKET NUMBER: CIT-2001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (207) 363-0558
; TELEFAX: (207) 363-0528
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6395 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 115..5385
US-09-228-317-1

Query Match 70.3%; Score 4435.4; DB 3; Length 6395;
Best Local Similarity 86.3%; Pred. No. 0;
Matches 4997; Conservative 0; Mismatches 721; Indels 69; Gaps 6;
Qy 590 ttcaggggcccgtcgttaaaagtgtgcctctgtctc-tccgaccggccacaggtttccgct 648

|||||
Db 8 TTCAGGGCCGTCGTAAGGTGTCGTCCTCGTCGCGCGGCCACACAGGTTTCCGCT 67
QY 649 tgcctctggccggggtcggcaactcagcggtcagtttccctcaagatggcgagcagg 708
Db 68 AGCTGGCGCGCGGGGTGCGGAACCTGCGGGGTTCGTTCCCTTAAAGATGCGCGACGAAG 127
QY 709 aggtcgaggtactagagagatgaaatcagcgcgaggtaccgccagccctcagcgtc 768
Db 128 AGATGGACGGCGGAGAGATGACGTCAGCCCGGAGCCTCCCTGGCCCGGACGCGGC 187
QY 769 tggcatcttggtgggacagcaagtgtattttatactgtcttcttgcatacttggcac 828
Db 188 CGGCATCGTGGGATCAGCAAGTTCATATCTACTGCTTCTTACATCATTTGGCAC 247
QY 829 aattggtgcagaaaaatttacttctgtcgaattggccccagacttggaaaaagcaggagaaa 888
Db 248 AATTAGTGCAGAAAATTTATTTTCTGAGATGGACCCAGATTTGGAAAAGCAAGAAGA 307
QY 889 gtacaaaatgtcaattaccctccactggaatggtacttatttggagaagatcacagata 948
Db 308 GTGTACAGATGTCAATACTCAGTCTTGGAGTGGTACTTATTTGGAGAGGATCCGGATA 367
QY 949 ttgtcttagaagaattgaagcacagtgagcatttcagcttgggtggaggggttttcaaaa 1008
Db 368 TTTGCTTAGAGAAATTAACACACAGTGGCGGTCCAGTGTGTGGGAAGTTTTCAAA 427
QY 1009 gtggagagacaacctattcttcaggaggtgtgcaattgtatcccaactgtgtactctgta 1068
Db 428 GTGGAGAAAACAATATCTCTGTAGGATGTGCAATTTGATCCCAACGTGTGCTCTGTA 487
QY 1069 tggactgtccaggacagtggttcaataaaaaatcactgttcaagaatgcatacttctactg 1128
Db 488 TGGACTGTCTCCAAAAGTAGTGTTCATAAAAACCATCTGTTCAAGATGCATCTTCTACTG 547
QY 1129 gagaggggtctgtgactgtggagacacagggcagtgaaactggcccttttgggtgtaa 1188
Db 548 GAGGGGCTTCTGTGACTGTGGAGACACAGAGCGTGGAAACTGGCCCTTTTGTGTGG 607
QY 1189 atcatgaacctggaagagcggtactataaaagaaattcacgctgtccgttgaatgaag 1248
Db 608 ATCAGGAGCTGGAAGAGCAGGTACTACAAAAGAGAGCTTACATTTGCCATTTGAATGAAG 667
QY 1249 agtgaattgtccagccagaaaaatttcttctcagtgataaaaaatgtcgtagaaaaatga 1308
Db 668 AGGTGATTTGCTCAAGCCAGAGAGATATTCCTTCGTTGATAAAAATACATTTAGAAAATGA 727
QY 1309 ctatatgggaagaaaaaagactgcctcctgaactccagataaaggagaaaaaatgaaa 1368
Db 728 CTATATGGGAAGAAAGAAAGTAATTCCTCTGAACTGCAGATGAAGGGAGAAAATGAAC 787
QY 1369 gatactattgtctcttttcaatgatgaacacatttcataatgaccacgctcatatacagcc 1428
Db 788 GATACATTTGTGTCTTTTCAACGATGAGCACCATTGCTATGATCATGTGATCTACAGTC 847
QY 1429 tacaagagctcttgactgtgactcgcagagcccgagttgcataccactgcccattgaca 1488
Db 848 TGCAGAGAGCTCTAGATTGCGAGCTTCAGAGGCGACAGCTGCACACGACTGCCATCGACA 907
QY 1489 aagagggctcgggctgttaagcgggagcttatgctgtgtgccaggaagcaaaagaaag 1548
Db 908 AAGAGGGTCCCGGGCTGTCAAGAGCAGGTGTATATGTCACCTTGGCAGGAAGCAAGGAGG 967
QY 1549 atataagagctattcagaaaaatgtctcacaactccacttcattgtagaaagtattacact 1608
Db 968 ATATAGAGAGTCACTCAGAGAACGTCTCTCAGACCCCTCCATGTGGAGTGTGCTGACT 1027
QY 1609 cagagattatggctcatcagaaaaattgtcttggcttgggttctcgtgatgacaaaaatta 1668
Db 1028 CCGTGGTTATGGCTCACCAGAAATTCGCTCTGCGCCTTGGCTCTCGATGAACAAAATTA 1087
QY 1669 tgagctattcaagtactttaggcagatcttttggccaagcatgctttagaagaagaaactg 1728
|||||

Db 1088 TGAGCTATTCAAGTGACTTTTAGACAGATATTTTGGCCAGGCGCTGCCGTGTAAGAAACCTG 1147
QY 1729 actcggagaataccctgtctcataaagcaggttaatagtttgggagtcgaaagctttataaag 1788
Db 1148 GCTCTGAAAAATCCCTGTCTTATAAGCAGACTAATGTCTTTGGGATGCAAAACTTTATAAAG 1207
QY 1789 gtgcccgaagatccctcattgaattgatctcagcagtttttttttggagatggaataca 1848
Db 1208 GTGCCCGTAAGATCCTTCATGAATTCATCTTTAGTAGTTTTTTTATTTGGAGATGAATACA 1267
QY 1849 aaaaactcttctgtatggaaatttgtgaaattattataaaacaactgcagaaagaatatata 1908
Db 1268 AAAAACTCTTCTATGGAATTTGTGAAGTATTATTAACAACATGCAGAAAAGATACATCA 1327
QY 1909 gtgatgataatgacaggaagatatctctataactgcactttcagtcagatgtttactgttc 1968
Db 1328 GCGACGACACAGAGAAAGCATCTCCATAACCGCCCTGTCGTCAGATGCTCACCGTCC 1387
QY 1969 ctactctggctcagcatcttattgaagacagaaatgttctctgtcattactcgaactc 2028
Db 1388 CGACCTTGGCCCGGCATCTTATTTGAAGACGAGAATGTATTTCTGTCAATTACTGAAACGC 1447
QY 2029 tgtagaagatttacctgagttacttgacaggaacaataaataaactccagggttata 2088
Db 1448 TGCTAGAAGTTTACTTGAATACTTGGACAGGAAACAATAAATCAACTTCCAGGGTTATA 1507
QY 2089 gccaggacaaattgggaagagtatatgcagtaatatgtgacctaaagtatatctcgtatca 2148
Db 1508 GCCAGGACAAACTGGGAAGAGTCTACCGCAGTTATATGTACCTTAAAGTATATCTGATTA 1567
QY 2149 gcaaacccacaatatggacagaaaagattaaagaatgcagttcccttgaaggttttcgatctt 2208
Db 1568 GCAAGCCTGTCTATATGGACAGAACCATTAAGACGCGAGTTCCTTGGGAAGGTTTCCGGTCTT 1627
QY 2209 ttttgaagattctacctgtatgcaggggaatggaagaaatccgaagacaggttgggcaac 2268
Db 1628 TTCTGAAGATCTTACCCTGTATGCAGGGAATGGAAAGAAATCAAGAACAACTTGGACAC 1687
QY 2269 acattgaagtggatcctgattgggaggtgccattgtctatacagatgcgaattgaagaata 2328
Db 1688 ACATTGAAGTGGACCTGACTGGAGGCTGCCATCGCTATACAGATGCAACTAAAGAATA 1747
QY 2329 ttttactcatgtccaaagagt 2388
Db 1748 TTTTCTCATGTTCGAAGAGT 1807
QY 2389 aagaatgcacaaaactgtgatgggtgcagtcaccagtttcatatctagttagcaaacag 2448
Db 1808 AAGAATGTCACAAAAGCTGTAATGAGTGCAGTACAAAATTTTCAATGTCCAGTACCACAGAC 1867
QY 2449 lagtaacatcgtgtggacatagtttggaaacaaaagtcctcacagatlatctgagatcttg 2508
Db 1868 TAGTCAATTGTGCGGTGCATAGCTCGGAACCAAACTCTTACAAAGTGTCTGAGGACCTTG 1927
QY 2509 taagcatatctgtccactctcttagacccttgcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2568
Db 1928 TAAGCATACACTGCCACTCTCTAGAACACTTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1987
QY 2569 tgggtgtgttttcaagactgcataaattgtgtcttttggagactttcagtagaggtac 2628
Db 1988 TAGTGTCTATTTCAGAGCTGCATGAATTTGTGCTTTTTCAGACGCTTTCAGTGTGAGGTCC 2047
QY 2629 lagtgaatatcttcaqgt 2688
Db 2048 TGGTGGAGTACCCGCTGCGCTGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2107
QY 2689 gaagaaatggagctgtctcttattagccaggtgttttattaccaagatgttaagtgcagag 2748
Db 2108 GAAGAAACGGGCTCTCACTCATCAGCCAGGTTTCTTATTATCAAGATGTTAAATGCAGGG 2167
QY 2749 aagaatgtatgataaagataatcatgcttcaagattggtgcattttaaagtatccca 2808
Db 2168 AGGAAATGTACGATAAAGATATCATGCTTCTCAGATTGGAGCATCTATAAATGGATCCCA 2227
|||||

Qy 2809 ataagctctgtactgtacttcagaggtatgaacttgcagagctttttaacaagacca 2868
Db 2228 ACAAGTCTTGTACTGTGTTACTTCAGAGATATGAATCTTACTGATGCTTTTAAACAAGACCA 2287
Qy 2869 tatctacaaaagaccagatttgattaaacaataaatacacactaatagaagaatgtcttc 2928
Db 2288 TATCCACAAAAGACCAGATTGTATTAAACAGTATTAATACATTAATAGAGAATAATGCTTC 2347
Qy 2929 aggtctcatctatatattggtgagcgttatgtacctggagtg9ggaatgtgacaaag 2988
Db 2348 AGGTCTCATCTATATTGTGGGAGAACGTTATGTACCTGGAGTGGAAATGTTACACAG 2407
Qy 2989 aagagtcacaatgagagaaatcatctacttgcgttttgattgaaccccatgcaacacagtg 3048
Db 2408 AGGAGGTTTATATGAGAGAGATTACTCTACTTACTTGTGATTGAGCCCATGCCACACAGTG 2467
Qy 3049 ccattgccaataattactcagaaatgaaataaagaactggcttagaagaatgtcataa 3108
Db 2468 CCATGCCCAAAACCTACTTGAGACGAAATTAATGAANAACCTGGCTTAGAGAAATGTCATAA 2527
Qy 3109 acaaagtgccacatttaagaacaccagtggtatcaggcccatggagtttatgaactaaaag 3168
Db 2528 ACAAAAGTGGCCACATTTAAGAAACAGGTGTGTGGGCCATGGAGTTTATGAATTGAAG 2587
Qy 3169 atgaatcactgaagacttcaatatgtacttttatacttccaaaccccgcatagca 3228
Db 2588 ATGAATCTACTGAAGACATTTCAATATGTACTTTTACCATTTATTTCTAAAACACACGATAGCA 2647
Qy 3229 aggtgaaacatatcagaagaaaagagaaaacaagaacaaaagatgaagcattgcgc 3288
Db 2648 AGGCTGAACATATGCAAGAAGAAAGGAGAAACAAAGAAATAAAGATGAAGCATTTGCCGC 2707
Qy 3289 caccaccacctcctgaaattctgcctgctttcagcaaatgattaaaccttccaactgtg 3348
Db 2708 GCCCACTCTCCAGAGTCTGCCCTGCTTTCAGCAAAAGTAGTCAACCTGCTCAGCTGTG 2767
Qy 3349 atatcatgtgtacattctcaggaaccgtattttagcgggcaatagacacagattctaaat 3408
Db 2768 ATGTATGATATATACATCTCTCAGGACCATCTTTGAGCGGCAGCTGGACACGGAGTCTAAATC 2827
Qy 3409 tctgacgaagggatgtcctcaaaatggcttttcataattctggcattgggtttactagaag 3468
Db 2828 TTGTGACAGAGAGGATGTGTCAGATGGGGTTCCTATATTTGGACATGGGCTTGTGGGAAG 2887
Qy 3469 aagaacacagcttcaaaaagctcctgagaagaagtaacatttgacttttatcataagg 3528
Db 2888 AGAACGACGCTTCAGAAAGCTCCTGAAGAGAGAGTGGCTTTTGACNTTTTACCATAAG 2947
Qy 3529 ctccaagattgggaagtccagccat-----gaatatacaaaatgcttttggaaaaac 3579
Db 2948 CTTCAGATTGGGAAGTTCAGCCATGAATGCTCAGAAATATACAAATGCTCTTTGGAAAGAC 3007
Qy 3580 tcaaaaggaattcccccagtttagaagccagaagaacatgataacgttggtacttcagatgt 3639
Db 3008 TCAAAAGGATCCCCAAATTTAGAAGGCCCAAGAGACATGATAACATGGATFACCTCCAGATGT 3067
Qy 3640 ttgcacagtgagcattaaagaaaaatctgttttaattgtagcaaacacacatcaggtat 3699
Db 3068 TTGCACAGTGAAGCGATTAAAGAAAAATCTGTGTTAGTTGGCAACCACTTTCAGGAC 3127
Qy 3700 cggaactlattaagaatgatgagattactcatgataaagaagaaagcagaacgaaaaagaa 3759
Db 3128 TGGAGTGCAATTAAGAGTGAGGAGATTACTCATGATTAAGAAAGGACGAGAAGCAAGAA 3187
Qy 3760 aagctgaagctgtaggctacatccaggaagatcatcaggtcagatgtcgtccttacaga 3819
Db 3188 AAGCTGAGGCCCTAGGCTTCATCGCAGAAGATCATGCCCCAGATGTCTGCTTTACAGA 3247
Qy 3820 aaaaacttaaggaaactcaaaactcatgtatgacaatacatcagaanaatgcctgggaag 3879
Db 3248 AAAACTTATTGAAACCCCAAAACTCATGTATGATAATACGTCAGAAAGTAACAGGGAAGG 3307

Qy 3880 aagattccattatggaggaagagcaccacagctcagctcagctactactctagaaattgctt 3939
Db 3308 AAGACTCCATTTATGGAGGAAGAGAGACACCTCAGCAGTCAGTCAGTGTAGCCCTCTAGAAATGCTC 3367
Qy 3940 tgggtctctaaacgggtccatctgttactgaaagaggtgctgacgtgcatcctcttgcc 3999
Db 3368 TGGGCCCTTAAACGGGGCCGGCTGTTACCGAAAAGAGAGGTGCTCAGCTGCATCCTCTGCC 3427
Qy 4000 aagaagaacagaggtgaaaaataatgccattggtattatcgctgctgctgtccaga 4059
Db 3428 AAGAAGAACAAGAGGTAAAACTAGAAAATATGCCATGGTATTGTCTACGATGTGTGCAGA 3487
Qy 4060 aatctactactcttaacccacgacacaggggaaaccccatagaaactcaggaagacccatg 4119
Db 3488 AATCCACCGCCCTTAAACCCAGCACAGAGGGAAGCCTGTGGACCCTTATAGGGAAACACTGG 3547
Qy 4120 accacttttcatggtccagacttggcatatgaaacttatcagaggaagctgtggtcatg 4179
Db 3548 ACCCTCTTTTCATGATGATCCAGACTTGGCATATGGAACCTTATACAGGAAGCTGTGTCATG 3607
Qy 4180 taatgcacgagtgctgtgcgagaagtattttgaagctgtacagctgagctctcagcagc 4239
Db 3608 TAATGTCATGCAGTGTGCTGCAGAAAGTATTTTGAAGCTGTGCAGCTGAGCTCGCAGCAGC 3667
Qy 4240 gaattcatgttgacotttttgaacttggaaagtggagaataatcttggccctcttggaaat 4299
Db 3668 GCATTCACCTTAGACCTGTTTGACCTGGAGAGCGCGAGTAGCTACCTATGCCCGCTCTGCAAGT 3727
Qy 4300 ctctgtcaactactgtgacctccattatctcttgaacctcaaaagataaaacagtga 4359
Db 3728 CTCTCTGCAACACTGTCTATCCCCCATCATCTCTTTCAGCCGCGCAGACAGATCAACAGTGA 3787
Qy 4360 atgcagatgctcttgcctcaacttttaccctggcagctgtaagagagaacac-----caatctctatttct 4476
Db 3788 ATGCGGAGGCTCTTCTCAACTTTTGACCTTGGCCGGTGGATACACACTGTCTCTGCCA 3847
Qy 4420 gaatacaggttataataataagaacatgctaaagagagaacac-----caatctctatttct 4476
Db 3848 GAATATCGGTTTATAATATTAAGCATGTCTAAAGGAGAAAGCCCAAGCAGTAGTTCTCTCTGT 3907
Qy 4477 ttaatacaagaaatggagattctacttggagttccattccattccactcctgagtttggcgtt 4536
Db 3908 TTAATCAAGAAATGGGGGATTCAACTTTTGAGTTTCAATTCATCTCCTGAGTTTGGAGTTC 3967
Qy 4537 agctcttcgattaaatattcaaaatagcatcaagaaaaatggttattctcttcttggcacaacaa 4596
Db 3968 AGCTTTCGGTGAATAATTCAAAATAGTATCAAGGAAATGGTCATTCTCTTGSCCAACAACAA 4027
Qy 4597 ttatgaattgattgaaagtgcacactgatgaaagggatcctcaggtccccaatgctga 4656
Db 4028 TTTTACAGAAATTTGGCCTGAAAGTGCTCTCTGATGAACCTAGACCCACGAGTGCCCATGTATGA 4087
Qy 4657 cctggagacactgcgtcttcaactatccaggcaattgaaaatctattggagatgaagaa 4716
Db 4088 CCTGGAGACAGTGTGGCTTACCATCCAGGCAATTCGAAAACCTTGTGGAGATGAAGGAA 4147
Qy 4717 aacctctgttggagcacttcaaaataggcagcataatggtctgaaagcattaatgcaat 4776
Db 4148 AACCTCTATTGGAGCACTTCAAAATAGACAGCATACACGGCTCTGAAGCGCTATATGCAAT 4207
Qy 4777 ttgcagttgcacagaggaattacctgtcctcaggtcctgatacagaacacatctggttcgtc 4836
Db 4208 TTGCAGTTGCACAGAGGGCTACCTGCCCTCAGGTCTCTGATACACAAACATCTGGCTGGC 4267
Qy 4837 ttctatcagttgttcttcttaacataaaaaatcagaagatacaccatgctctctgctatag 4896
Db 4268 TCCTGTCAAGTTATTCTTCTTAACCTGAATCAGAAAAATACACAGGCTTCTGTCTGTGG 4327
Qy 4897 atctgttcaatgttgggtgggtgctgttagcattcccatcctcttattggtgagatgacc 4956
Db 4328 ATCTCTTCCATGTTCTGGTGGCGCAGCTTTAGGCTTCCCATCTCTGTATTGTTGGATGACA 4387
Qy 4957 ctgttgatctgcagccttcttcagtttagttcttcttaacacacacttttatctcttccatt 5016


```
Db 1 ATGGATCCCAACAGTCTTGTTACTGTTACTGATGAGGTATGAACCTGCCAGGCTTTT 60
QY 2859 acaagaccatctacaaaaccaggatttgattaaacaataataacactaatagaa 2918
Db 61 AACAGACCATATCTACAAAGACAGGATTTGATTAAACAATAATAACATAATAGAA 120
QY 2919 gaaatgcttcaggtcctcatctatatttggtgagcgcttctacccggagtggaat 2978
Db 121 GAAATGCTTCAGTCTCATCTATATTGTGGTGAGCGTTATGCTACCTGGAGTGGGAAAT 180
QY 2979 gtgaccaagaaggtcacatgagagaaatcattcactcttcttgattgaacccatg 3038
Db 181 GTGACCAAGAAGAGGTCAATGAGAGAAATCATTCACCTTGCTTTGCAATTGAACCCATG 240
QY 3039 ccacacagtgccatgcccataattacctgagatgaaataaataagaaactggcttagag 3098
Db 241 CCACACAGTGCCATTGCCNAATTTACCTGAGATGAAATAATGAAACTGGCTTAGAG 300
QY 3099 aatgtcataaacaagtggccacatttaagaaccagggtgtatcaggccatggagttat 3158
Db 301 AATGTCATAAACAAGTGCCACATTTAAGAAACAGGTTGATCAGGCCATGGAGTTTAT 360
QY 3159 gaactaaagatgaatcactgaaagacttcaatgtacttttatacttactccaaaacc 3218
Db 361 GAACTAAAGATGAATCACTGAAGACTTCAATATGTACTTTTATCATTTACTCCAAACC 420
QY 3219 cagcatagcaagctgaacatgcagaaagaaagagaaacaaagaaacaaagatgaa 3278
Db 421 CAGCATAGCAAGCTGAACATATGCAGAAAGAGAGAAACAAAGAAACAAAGATGAA 480
QY 3279 gcattgcgcaccaccacctcctgaattcgcctctgttcagcaaatgattacacctt 3338
Db 481 GCATTGCGCCACCACCACCTCTCTGAATTCGCTCTGCTTTTCAGCAAGTGATTAACCTT 540
QY 3339 ctcaactgtgatatacatgatgtacattctcaggaccgttatcttgagcgggcaatagacaca 3398
Db 541 CTCAACTGTGATATCATGTATGATCTCTCAGGACCGTATTTCAGCGGGCAATAAACACA 600
QY 3399 gattctaactgtgacgaagaggtgctccaaatggcttttcattcttctgacattggt 3458
Db 601 GATTCTAACTTGTGACCGGAAGGATGCTCCAAATGGCTTTTCATATTCTGCGATTGGGT 660
QY 3459 ttactagaagaagcaacagcttcaaaaagctccctgaagaagaagtaacatttgacttt 3518
Db 661 TTACTAGAGAGAAGCAACAGCTTCAAAAAGCTCTCGAAGAGAAGTACATTTGACTTT 720
QY 3519 tatcataaggtctcaagattgggaagttcagccatgaaatacaaatgcttttggaaaaa 3578
Db 721 TATCATAGGCTTCAAGATTGGGAAGTTCAGCCATGAATATACAAATGCTTTTGGAAAAA 780
QY 3579 ctcaagaagaattcccagttagaagccagaaggacatgataacgtggatacttcagatg 3638
Db 781 CTCAAAGGAATTTCCCGAGTTAGAAGGCCAGAAGGACATGATAACGTGGTACTTTCAGATG 840
QY 3639 ttgacacagtgaaagcgttaagaaataatctgtttaattgtagcaaccacatcaga 3698
Db 841 TTGACACAGTAGAGCGATTAGAGAAAAATCTTGTTTAATTGTAGCAACCATCAGGA 900
QY 3699 tcggaatctattaagaatgatgattactcatgataaagaaaaagcagaacgaaaaaga 3758
Db 901 TCGGAATCTATTAAAGATGATGAGATTACTCATGATATAAGAAAAAGCAGAACGAAAAAGA 960
QY 3759 aaagctgaagctgctaggctacatgcagcagaagatcagc 3799
Db 961 AAAGCTGAAGCTGCTAGGCTTCATGCCCAAGAGATCATGGC 1001
```

RESULT 4

US-09-228-317-2

; Sequence 2, Application US/09228317

; Patent No. 6159732

; GENERAL INFORMATION:

```
; APPLICANT: Varshavsky, Alexander
; TITLE OF INVENTION: NUCLEIC ACID ENCODING MAMMALIAN UBRL
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kevin M. Farrell, P.C.
; STREET: P.O. Box 999
; CITY: York Harbor
; STATE: ME
; COUNTRY: US
; ZIP: 03911
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/228,317
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Farrell, Kevin M.
; REGISTRATION NUMBER: 35,505
; REFERENCE/DOCKET NUMBER: CIT-2001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (207) 363-0558
; TELEFAX: (207) 363-0528
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1001 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..999
; US-09-228-317-2
```

```
Query Match 15.8%; Score 996.2; DB 3; Length 1001;
Best Local Similarity 99.7%; Pred. No. 7.6e-279;
Matches 998; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2799 atgagtcaccaataagttctgttactgtacttcagaggtatgaacttgcgcagggctttt 2858
Db 1 ATGATCCCCAACAAAGTCTTGTTACTGTTACTGTTACTGAGGTATGAACTTGGCAGGCTTTT 60
QY 2859 acaagaccatctacaaaaccaggatttgattaaacaataataacactaatagaa 2918
Db 61 AACAGACCATATCTACAAAGACAGGATTTGATTAAACAATAATAACATAATAGAA 120
QY 2919 gaaatgcttcaggtcctcatctatatttggtgagcgcttctacccggagtggaat 2978
Db 121 GAAATGCTTCAGTCTCATCTATATTGTGGTGAGCGTTATGCTACCTGGAGTGGGAAAT 180
QY 2979 gtgaccaagaaggtcacatgagagaaatcattcactcttcttgattgaacccatg 3038
Db 181 GTGACCAAGAAGAGGTCAATGAGAGAAATCATTCACCTTGCTTTGCAATTGAACCCATG 240
QY 3039 ccacacagtgccatgcccataattacctgagatgaaataaataagaaactggcttagag 3098
Db 241 CCACACAGTGCCATTGCCNAATTTACCTGAGATGAAATAATGAAACTGGCTTAGAG 300
QY 3099 aatgtcataaacaagtggccacatttaagaaccagggtgtatcagcgcattagag 3158
Db 301 AATGTCATAAACAAGTGGCCACATTTAAGAAACAGGTTGATCAGGCCATGGAGTTTAT 360
QY 3159 gaactaaagatgaatcactgaaagacttcaatgtacttttatacttactccaaaacc 3218
Db 361 GAACTAAAGATGAATCACTGAAGACTTCAATATGTACTTTTATCATTTACTCCAAACC 420
QY 3219 cagcatagcaagctgaacatgcagaaagaaagagaaacaaagaaacaaagatgaa 3278
```



```

; TELEFAX: 650-327-3231
;
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 72928 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
US-09-009-913-1

Query Match 0.7%; Score 44.2; DB 3; Length 72928;
Best Local Similarity 86.0%; Pred. No. 0.18;
Matches 49; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 303 cctcggaacacacattctactttctgtctatgaatttgactactctagctgcat 359
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
Db 19091 CCGCGCAACACCATCTACTTTCTGCTCTATGAATTTGACCACTCTAAGTAGCT 19035

RESULT 9
US-08-991-789A-236/C
; Sequence 236, Application US/08991789A
; Patent No. 6225054
; GENERAL INFORMATION:
; APPLICANT: Fridakis, Tony N.
; Smith, John M.
; Reed, Steven G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF BREAST CANCER
; NUMBER OF SEQUENCES: 292
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed IP Law Group
; STREET: 701 Fifth Avenue, Suite 6300
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/991,789A
; FILING DATE: 11-Dec-1997
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Potter, Jane E. R.
; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 210121.419C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 236:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 325 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; SEQUENCE DESCRIPTION: SEQ ID NO: 236:
US-08-991-789A-236

Query Match 0.7%; Score 42.2; DB 4; Length 325;
Best Local Similarity 79.4%; Pred. No. 0.021;
Matches 50; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 298 acacactctggaacaccattctactttctgtctatgaatttgactactctagctg 357

```


US-08-299-953-1

Query Match 0.6%; Score 39.8; DB 1; Length 2861;
Best Local Similarity 46.5%; Pred. No. 0.42;
Matches 128; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

Qy 3607 agaagacatgataaacgtgatcattcgatcttggacacagtggaacgcgattaaagagaaa 3666
| | | | |
Db 1290 AAAAAGAACGGAGGAGCTTAATATTTTGTAGATTTTACACGTAATTAAAAAAATATATCAA 1231
| | | | |
Qy 3667 aactctgttaattgttagcaaccacacatcaggatcggaatctatttaaagaatcgatgagatta 3726
| | | | |
Db 1230 AAAAATAATCTTTATATATTAATAAAAATGGAGAGATAAATTTATATTAATTAATAAAAAAAG 1171
| | | | |
Qy 3727 ctcatgataaagaaaaagcacgaacgaaaaagaaagctgaagctgctgagctacatcgcc 3786
| | | | |
Db 1170 ATAATAATAAAATCTAGAGTTATAATAAAAAACTAATAATTAATTCITTTTAAACATTGTA 1111
| | | | |
Qy 3787 agaagatcattgcttcagatgtcgccttacagaaaaaacttcattgaaactcataaactca 3846
| | | | |
Db 1110 AATGATTTATATATCATATAATTTTTTCAAAACAACCAATTAATAAAAAATGATAGGG 1051
| | | | |
Qy 3847 tgmtacacaatcacatcagaataatgocctgggaaagaa 3881
| | | | |
Db 1050 AGTATTATCATATGTCAGAAATATTATAAAGAA 1016
| | | | |

RESULT 13
US-08-459-415-1/c
; Sequence 1, Application US/08459415
; Patent No. 5744334
; GENERAL INFORMATION:
; APPLICANT: Dobres, Michael S. and Mandaci, Sevnur
; TITLE OF INVENTION: A Plant Promoter Useful for Directing the
; NUMBER OF SEQUENCES: 4 Expression of Foreign Proteins to the Plant Epidermis
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5744334ris
; STREET: One Liberty Place 46th. Floor
; CITY: Philadelphia
; STATE: PA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,415
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/299,953
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Beardell, Lori Y.
; REGISTRATION NUMBER: 34,293
; REFERENCE/DOCKET NUMBER: NOVA-0003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-564-8960
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2861 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-459-415-1

Query Match 0.6%; Score 39.8; DB 1; Length 2861;
Best Local Similarity 46.5%; Pred. No. 0.42; 147; Indels 0; Gaps 0;
Matches 128; Conservative 0; Mismatches 147; Indels 0; Gaps 0;
QY 3607 agaagacatgataacgctggtgacttcagatgtttgacacagtgagcgattaaagagaaa 3666
DB 1290 AAAAAAGACGGAGGAGTAAATATTTTAGATTTTACACGATTTAAAAAATATATCAAA 1231
QY 3667 aatctgtttaattgtagcaaccacatcaggatcggaatctatttaagaatgagatta 3726
DB 1230 AAAAATATCTTTATATATTAATAAATGGAGAAGATAAATTTATATAATTAATAAAG 1171
QY 3727 ctcatgataaagaaacgagaaacgagaaacgagaaacgagaaacgagaaacgagaa 3786
DB 1170 ATAATATAAAATCTAGAGTTTATATAAATAAATAAATAAATAAATAAATAAATAA 1111
QY 3787 agaagatcgtcgtcagatgctgccttacagaaaaacacacacacacacacacacac 3846
DB 1110 AAATGATTATATATATATAAATTTTTCARAAACACCATATAAATAAATGATAGG 1051
QY 3847 tgtatgacatcacatcagaaatgctgggaaagaa 3881
DB 1050 AGTATTATCATATGTCAGAAATTTATATAAAGAA 1016

RESULT 14
PCT-US95-11231-1/c
; Sequence 1, Application PC/TUS9511231
; GENERAL INFORMATION:
; APPLICANT: Dobres, Michael S. and Mandaci, Sevnur
; TITLE OF INVENTION: A Plant Promoter Useful for Directing the Expression
; TITLE OF INVENTION: of Foreign Proteins to the Plant Epidermis
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & Norris
; STREET: One Liberty Place 46th. Floor
; CITY: Philadelphia
; STATE: PA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/11231
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/299,953
; FILING DATE: September 2, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Beardell, Lori Y.
; REGISTRATION NUMBER: 34,293
; REFERENCE/DOCKET NUMBER: NOVA-0016
; TELEPHONE: 215-564-8960
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2861 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
PCT-US95-11231-1

Query Match 0.6%; Score 39.8; DB 5; Length 2861;
Best Local Similarity 46.5%; Pred. No. 0.42;

Matches 128; Conservative 0; Mismatches 147; Indels 0; Gaps 0;
QY 3607 agaagacatgataacgctggtgacttcagatgtttgacacagtgagcgattaaagagaaa 3666
DB 1290 AAAAAAGACGGAGGAGTAAATATTTTAGATTTTACACGATTTAAAAAATATATCAAA 1231
QY 3667 aatctgtttaattgtagcaaccacatcaggatcggaatctatttaagaatgagatta 3726
DB 1230 AAAAATATCTTTATATATTAATAAATGGAGAAGATAAATTTATATAATTAATAAAG 1171
QY 3727 ctcatgataaagaaacgagaaacgagaaacgagaaacgagaaacgagaaacgagaa 3786
DB 1170 ATAATATAAAATCTAGAGTTTATATAAATAAATAAATAAATAAATAAATAAATAA 1111
QY 3787 agaagatcgtcgtcagatgctgccttacagaaaaacacacacacacacacacacac 3846
DB 1110 AAATGATTATATATATATAAATTTTTCARAAACACCATATAAATAAATGATAGG 1051
QY 3847 tgtatgacatcacatcagaaatgctgggaaagaa 3881
DB 1050 AGTATTATCATATGTCAGAAATTTATATAAAGAA 1016

RESULT 15
US-08-299-953-2/c
; Sequence 2, Application US/08299953
; Patent No. 5646333
; GENERAL INFORMATION:
; APPLICANT: Dobres, Michael S. and Mandaci, Sevnur
; TITLE OF INVENTION: A Plant Promoter Useful for Directing the
; TITLE OF INVENTION: Expression of Foreign Proteins to the Plant Epidermis
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5646333ris
; STREET: One Liberty Place 46th. Floor
; CITY: Philadelphia
; STATE: PA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/299,953
; FILING DATE: Herewith
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Beardell, Lori Y.
; REGISTRATION NUMBER: 34,293
; REFERENCE/DOCKET NUMBER: NOVA-0003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-564-8960
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3881 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-299-953-2

Query Match 0.6%; Score 39.8; DB 1; Length 3881;
Best Local Similarity 46.5%; Pred. No. 0.51;
Matches 128; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

QY 3607 agaagacatgataacgctggtgacttcagatgtttgacacagtgagcgattaaagagaaa 3666
DB 1290 AAAAAAGACGGAGGAGTAAATATTTTAGATTTTACACGATTTAAAAAATATATCAAA 1231

```

Qy 3667 aatctgtttaatttagcaaccacacatcaggatcggaatcttatttaagaatgatgagatta 3726
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1230 AAAAATATCTTTATATATATAATAAATGAGAGATAAATTTATATAATTAATAAAG 1171
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 3727 ccatgataaagaaagcagaaagaaagaaagcgaagctgtaggctacatcgcc 3786
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1170 ATATAATAAATTTCTAGAGTTATATAAATAAATAAATAAATAAATAAATAAATAA 1111
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 3787 agaagatcatgggctcagatgtctgccttacagaaataaacttcattgaaactcataaactca 3846
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1110 AAATGATTTATATATATATAAATTTTTCAAAACAACCATATAATAAATAATGATAGG 1051
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 3847 t9tatgacaatacatcagaaatgcctg9gaaagaa 3881
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1050 AGTATTATCATATGTCAGAAATTTATTATAAAGAA 1016
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Search completed: March 1, 2002, 16:12:22
 Job time: 12692 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 1, 2002, 12:32:35 ; Search time 2649.68 Seconds
(without alignments)
25582.097 Million cell updates/sec

Title: US-09-724-126A-1

Perfect score: 6308

Sequence: 1 gccagaattgcgcagagg.....aattttgtatttggtgtttt 6308

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: gb_est1:*
11: gb_est2:*
12: gb_hic:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	674.6	10.7	682	10	AI929033
2	656.6	10.4	756	11	BG534574
3	637.4	10.1	641	10	AI361043
4	627.2	9.9	797	11	BG862813
5	573.2	9.1	782	11	BI086469
6	557	8.8	565	11	BF063405
7	512.2	8.1	606	10	BE589438
8	479.8	7.6	624	11	BG219270
9	477.8	7.6	481	10	AI192195
10	473.4	7.5	583	10	AW971391
11	471.6	7.5	478	10	AA401319
12	454	7.2	522	10	AW291190

13	449	7.1	522	11	BG382648
14	448	7.1	520	11	BG382624
c 15	410.8	6.5	486	10	BE650873
16	354.8	5.6	925	10	BE573578
c 17	352	5.6	374	10	AA400279
18	351.4	5.6	435	10	AW489271
19	330	5.2	394	11	BE774117
c 20	318	5.0	481	10	AI187306
21	307.2	4.9	397	10	AW311960
22	301.8	4.8	504	10	AI980640
23	266.2	4.2	327	10	BE654236
24	265.2	4.2	523	11	BG797647
c 25	263.4	4.2	300	11	Z17892
26	252.6	4.0	1079	11	BG292980
c 27	249	3.9	562	10	AI504731
28	243.6	3.9	372	11	BF542537
29	242.8	3.8	505	11	BG364916
30	241	3.8	325	10	AI615529
31	237	3.8	291	10	AV225341
c 32	235.8	3.7	396	11	BE930958
33	232.8	3.7	256	10	AI921294
c 34	231.4	3.7	278	10	BE077143
35	230.6	3.7	323	11	H33916
c 36	223.6	3.5	745	10	AW976158
37	222	3.5	555	11	BE873236
38	217	3.4	685	10	AI693180
39	209	3.3	313	10	AV168252
40	208.6	3.3	632	11	BG625558
41	206.8	3.3	221	10	AA507138
42	204	3.2	1148	11	BF164318
c 43	201	3.2	620	10	AW702134
44	200.4	3.2	435	11	BE930879
c 45	200	3.2	711	10	AI646734

ALIGNMENTS

RESULT 1

AI929033

LOCUS

DEFINITION

IMAGE:2519538 5' similar to TR:070481 O70481 UBIQUITIN-PROTEIN

LIGASE E3 COMPONENT N-RECOGNIN ;, mRNA sequence.

ACCESSION

AI929033

VERSION

AI929033.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

AUTHORS

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (1997)

Other ESTs: au64c10.xl

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Possible reversed clone: similarity on wrong strand

Seq primer: -40RP from Gibco

High quality sequence stop: 450.

Location/Qualifiers

1..682

/organism="Homo sapiens"

23-AUG-1999

EST

682 bp

mRNA

00004 Homo sapiens cDNA clone

IMAGE:2519538 5' similar to TR:070481 O70481 UBIQUITIN-PROTEIN

LIGASE E3 COMPONENT N-RECOGNIN ;, mRNA sequence.

AI929033

GI:5664997

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 682)

Miller, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,

Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin

, J., Moore, B., Scheilberg, K., Steptoe, M., Tan, F., Theising, B.,

White, Y., Wylie, T., Waterston, R. and Wilson, R.

WashU-NCI human EST Project

Unpublished (1997)

Other ESTs: au64c10.xl

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Possible reversed clone: similarity on wrong strand

Seq primer: -40RP from Gibco

High quality sequence stop: 450.

Location/Qualifiers

1..682

/organism="Homo sapiens"

```

/db_xref="taxon:9606"
/clone="IMAGE:2519538"
/clone_lib="Schneider fetal brain 00004"
/sex="male"
/tissue_type="frontal lobe"
/dev_stage="5 months post-conception"
/lab_host="DH10B"
/notes="Organ: brain; Vector: pBluescript SK (Stratagene);
Site_1: Ssti; Site_2: XhoI; Double-stranded cDNA was
prepared from human fetal brain tissue. 5' and 3'
adaptors were used in cloning as follows: 5' adaptor
sequence:
5'-GAGAGAGAGAGAGCTCAAGATGCTTAAATAATTAATCCCCCCCCCC-3'
and 3' adaptor sequence:
5'-GAGAGAGAGACTCGAGTTTTTTTTTTTTTTT-3'. The library was
size-selected for >0.5 kb inserts and has an average
insert size estimated at 1.2 kb. This library was
constructed using the CAP-trapper method for full-length
enrichment and has not undergone amplification. Library
was constructed by Dr. Claudio Schneider (LNCIB-Area
Science Park, Trieste, Italy)."
BASE COUNT      178 a  155 c  166 g  182 t      1 others
ORIGIN

Query Match      10.7%; Score 674.6; DB 10; Length 682;
Best Local Similarity 99.3%; Pred. No. 3.9e-147;
Matches 677; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 5261 aactccgctgagaactgcataccaattctgcagaaggagctacagtcaactgtag 5320
DB 1  AACTCCGCGCTGAGAACTGCAATACCAATTCGCAAGAGAGTAGACGTGCACACTGTGATG 60

QY 5321 ctatctatcttacctaca aaattgttctgctctctccaggaatattgggatactgtaag 5380
DB 61 CTATCTATCTTTACCTACAAATTTGTTCTGCTCTCTCCAGGAATATTGGGATACCTGTAAG 120

QY 5381 gcccttgctccagagggtgtgtgcagatccttgccttactaaactgtttgaagc aaaaaa 5440
DB 121 GCCCTTGCTCCAGAGTGGTGTGCAGATPCCTGCCCTTACTAAACTGTTTGAAGCAAAAAA 180

QY 5441 caccgtgctcaggttacccatagaaaaaagattgttgaatagagcttctctgatactatag 5500
DB 181 CACCGTGTGAGGTACCCCTAGAAAAGAAATAGTTTGTATAGAGCTTCTGTATGACTATAG 240

QY 5501 ctgctctcgtaatcaagctctctcatcttcaggtgccacggctctgcagatgatgcgcgaaa 5560
DB 241 CTGCTCTCTGAATCAAGCTTCTCATTTTCAGGTGCCACGGTCTGCAGATGATGAGCGAAA 300

QY 5561 gcatactgctctgcctttctgtggggtatactatgttctcagaacatttgtctgcca 5620
DB 301 GCATCCTGCTCTGCGCTTTCTGTGGGGCTATACATATGTTCTCAGAACATTTGCTGGCCA 360

QY 5621 ggaattgtgaacggggaagggttgagctgtcatttttcacgcacttcactgtggagc 5680
DB 361 GGAATTTGTGAACGGGGAAGAGTTTGGAGCTTGCAATTTTTTTCAGGCACCTCACTGTGGAGC 420

QY 5681 cggagctgcattttcctaaaaatacagagaatgccagtggtgctctgggtgaaggtaaaagc 5740
DB 421 CGGAGTCTGCATTTTTCCTAAAAATCAGAGAATGCCAGTGGTCTCGTTGAAGGTAAAGC 480

QY 5741 cagaggctgtgccctatccagctccttacttgatgaaataggagaaacagaccctggcct 5800
DB 481 CAGAGGCTGTGCCCTATCCAGCTCCTTATCTGGATGAATATGGAGAAACAGACCCCTGGGCT 540

QY 5801 gaagaggggcaacccctctattatctcgtgagcgggtatcggaagctccatttggctctg 5860
DB 541 GAAGAGGGGCAACCCCTTCATTTATCTCGTGCAGCGGTATCGGAAGCTCCATTTGGTCTG 600

QY 5861 gcaacaaactgcattatagaagagattgctaggagccaaagacataatcagatgttatt 5920
DB 601 GCAACAACTGCAATTATAGAAGAGATTGCTAGAGCCAAAGAGACTAATCAGATGCTTATT 660

```

[illegible]

QY 3076 aaaaatgaactggcttagaagatgcataaaagtgccacatttaagaaaccag 3135
|||||
Db 241 AAAATATGAAGTGGCTTAGAGATGCTATAAACAAGTGGCCACATTTAAGAAACAG 300
QY 3136 gtatcagcccgatgagttatgaactaaaagatgaactcactgaaagacttcaatgt 3195
|||||
Db 301 GTGTATCAGCCCATGGAGTTATGAATAAAGATGAATCACTGAAAGACTTCAATATGT 360
QY 3196 actttatcattactcaaaaccagcatagcaagctgaacatatcagaagaagaaggga 3255
|||||
Db 361 ACTTTTATCATTTACTCAAAACCCAGCATAGCAAGGCTGAACATATGCAGAAGAAAGGA 420
QY 3256 gaaacaagaaaaaagaatgaag-----cattgcgcgcacacaccctccctgaattct 3309
|||||
Db 421 GAAACAAGAAAAACAAGATGAAGTTAAACATTTGCCGACACCACCTCCTGAATCT 480
QY 3310 gccctgcttcagcaagtgattcaactctcaactgtgat-catgatgtacattctc 3368
|||||
Db 481 GCCTGCTTTTACGAAGTGAATTAACCTTCAACTGTGATATCCATGATGTACATTCTC 540
QY 3369 aggaccgtattgagcgccaatagacagattctaaacttggaaccggaaggatgctc 3428
|||||
Db 541 AGGACCGTATTTAGCGGGCAATAGACAGATTCCTAACCTGTGATATCCATGATGTACATTCTC 600
QY 3429 caaatggctttcattctctggcattggttttactagaagaagcaaacagcttcaaaa 3488
|||||
Db 601 CAAATGGTTTCATATTTCTGGCATTTGGGTTTACTAGAAGAGAGAACAG-TTCAAAA 659
QY 3489 gctcctgaagaagaagtaacattgacttttatacagaagcttcaagattgggaagtcca 3548
|||||
Db 660 GCTCCTGAAGAAGTAACATTTGACTTTTATCATAAAGGCTTCAAGATTGGGAAGTTCA 719
QY 3549 gccatgaatatcaaatgcttttggaaaaactcaaaaggaat 3589
|||||
Db 720 G-CATGAATATACAA---TGTTTTGAAAAAATCAAAAGGATT 756

RESULT 3
A1361043/c
LOCUS
DEFINITION
qy03f11.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:2010957 3'
similar to TR:P91133 P91133 SIMILAR TOS. CEREBISIAE
UBIQUITIN-PROTEIN LIGASE E3 COMPONENT SP:P19812. ;, mRNA sequence.
A1361043
A1361043.1 GI:4112664
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 641)
NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/BTGP), Tumor Gene Index
unpublished (1998)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution Information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert length: 1083 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 453.
Location/Qualifiers
1. .641
/organism="Homo sapiens"

FEATURES
source

/db_xref="taxon:9606"
/clone="IMAGE:2010957"
/clone_lib="NCI_CGAP_Brn23"
/tissue_type="glioblastoma (pooled)"
/lab_host="DH10B"
/note="Organ: brain; Vector: p7T3D-Pac (Pharmacia) with a
modified polylinker; Site:1: Not I; Site:2: Eco RI; lstr
strand cDNA was primed with a Not I - oligo(dt) primer [5',
TGTTACCAATCTGAAGTGGAGCGCCGATATCTTTTTTTTTTTTTTTTTTTT
T 3']; Double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified p7T3 vector.
Library is normalized, and was constructed by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 174 a 155 c 135 g 175 t 2 others
ORIGIN
Query Match 10.1%; Score 637.4; DB 10; Length 641;
Best Local Similarity 99.5%; Pred. No. 2e-138;
Matches 638; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 5457 cctagaaaagaagaatgtttgatagacttctctgactatagctgcctcctgaatcaa 5516
Db 641 CCTAGAAAAGAATAGTTTGTAGAGCTTCTGTGATGACTATAGCTGCCTCTNGATCCA 582
QY 5517 gcttctcattcagtgccacggtctcgatgatgagcgaagcatcctgtcctctgc 5576
Db 581 GCTTCTCATNTCAGTGGCCACGGTCTCAGATGATGAGCGAAGCATCTGTCTCTGC 522
QY 5577 ctttctg9gggtatactatgttctcagaacatttgcctgcaggaaaattgtgaacgg 5636
Db 521 CTTTCTGTGGGGTATATCTATGTTCTCAGAACATTTGTCGACGAAATTTGTAACGG 462
QY 5637 gaagaggttgagcttgcatcttttccacacttccactgtgagcggagctgcatttc 5696
Db 461 GAAGAGGTTGAGGCTTGCAATTTTTCACGACATTCCTGTGGAGCGGAGTCTGCAFTTTC 402
QY 5697 ctaaaaatcagagaatgccgagtgctctggtgaaggttaaagccagagctgtgcctat 5756
Db 401 CTAAAAATCAGAGATGCCGAGTGGTCTGGTTGAAGGTAAAGCCAGAGCGCTGTGCCTAT 342
QY 5757 ccagctccttacttgatgaataatggagaacacagaccctggcctgaagaggggcaacccc 5816
Db 341 CCAGCTCCTTACTTGGATGAATATGAGAAACAGACACCCTGGCTGGAAGAGGGCAACCCC 282
QY 5817 ctctattatctctgagcggatcggaagctccatttggcttggaacaaacactgcatt 5876
Db 281 CTTTCAATTTATCTCTGTAGCGGTATCGGAAGCTTCCATTTGGTCTGGCAACACACATGCAAT 222
QY 5877 atagaagagattgctagagccaaagagactaatcagatgttatttgattcaactggcag 5936
Db 221 ATAGAAAGATTTGCTAGAGCCAAAGACTAATCAGATGTTATTGGAATCACTTGGGAG 162
QY 5937 ttactgtagctccaactctgcctcaagacaatcacaatgacgcagtagtaaggctg 5996
Db 161 TTACTGTGAGCTCCAACTCTGCCTCAAGACAATCACAATACGACACACTAGTAAGGCTG 102
QY 5997 attcaaatattggaagaacttctgagggctgggaagatttgagggtcttttgcctca 6056
Db 101 ATTCAAAATTTATGAAAACCTTCTGTAGGGCTGGGAAAGTATTGGAGGCTCTTTTGTCTCCA 42
QY 6057 tgcacaggttcacttacatacaataataatttcttaatgga 6097
|||||
Db 41 TGTCCAGGTTCACTTACATCAATAAAAAATATTTCTTAATGGA 1
RESULT 4
BG862813
LOCUS
DEFINITION
602799074F1 NIH_CGAP_Mam4 Mus musculus cDNA clone IMAGE:4934370 5',
mRNA sequence.
ACCESSION
BG862813

```

VERSION      BG862813.1  GI:14213351
KEYWORDS     EST.
SOURCE       house mouse.
ORGANISM     Mus musculus

REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE        NIH-MGC http://mgi.nci.nih.gov/.
JOURNAL      National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT      Unpublished (1999)
              Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth
              Ph.D.
              CDNA Library Preparation: Life Technologies, Inc.
              DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
              CDNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLAM10863 row: d column: 19
              High quality sequence stop: 690.
              Location/Qualifiers
                1..797
                  /organism="Mus musculus"
                  /db_xref="taxon:10090"
                  /clone="IMAGE:4934370"
                  /clone_lib="NIH_CGAP_Mam4"
                  /tissue_type="tumor, gross tissue"
                  /lab_host="DH10B"
                  /note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: NotI;
                  Site:2: SalI; Cloned unidirectionally. Primer: Oligo dt.
                  Average insert 2.5 kb. Library constructed by Life
                  Technologies, catalog # 12018-016. Investigators providing
                  samples: Lothar Hennighausen/Priscilla Furth, NIH
                  Reference for transgenic model: Li et al., Cell Growth and
                  Differentiation 7, 3-11 (1996). Note: This is a NCI_CGAP
                  Library."
BASE COUNT   194 a 177 c 207 g 219 t
ORIGIN
Query Match          9.9%; Score 627.2; DB 11; Length 797;
Best Local Similarity 87.6%; Pred. No. 5e-136;
Matches 697; Conservative 0; Mismatches 98; Indels 1; Gaps 1;

QY 5171 tcttggtggtattgtgggtcactgaagaatggcacccttatcttcgtgtgc 5230
DB 2 tccccggctgggtaccgtggtggtcctccctgaggaacgcacaccccttaccctccgtgtgc 61
QY 5231 tgcattgtttccactatttacttgggttaactccgcctgaggaactgcataccaattc 5290
DB 62 tgcactgtctttccactatttacttgggtgagtagctccgcctgaggaactgttgcctaattc 121
QY 5291 tgcagaagagagtagtcagtcactctgtagctatctatttacctacaaattgttctc 5350
DB 122 tgcctgagagagatcagtcgactctgtagctatctatttacctacaaattgttctc 181
QY 5351 gctctccagggaatttgggtactgaaggcccttgcctccagagtggtgtgcagatcc 5410
DB 182 gcttttccagggaatttgggtactgaaggcccttgcctccagagtggtgtgcagatcc 241
QY 5411 tgccttactaaactgttgaagcaaaaaaacacccgtggtcaggtacccttagaaaaagaaa 5470
DB 242 tgccttactaaactgttgaagcaaaaaaacacccgtggtcaggtacccttagaaaaagaaa 301
QY 5471 tagttgatagactcctgtagctatgactatgctgctcctgaatcaagcttcatttcag 5530
DB 302 tagttgatagactcctgtagctatgactatgctgctcctgaatcaagcttcatttcag 361
QY 5531 gtgccacggtctgcagatgatgagcgaagcagcattcctgctccttcttctgtgtgggc 5590
DB 362 gtgtccacgggtctgcagatgatgagcgaagcagcattcctgctccttcttctgtgtgggc 421

QY 5591 tatactatgttctcagaacaatttgcctgcccaggaataattgtgaacggggaagaggttgagc 5650
DB 422 CATCCCTGTGTCTCAACAACATCTGTTGCCAAGAAATAGTGAATGGGGAAGAGGTGGAGC 481
QY 5651 ttgcatttttcacgcacttcactgtgagccgagtcctgattttctctaaaaaacagaga 5710
DB 482 GTGCGTTTTTCATCGCTTCATTTGGTGTGTGGAGTCTGCAATTTTCCTAAAAATCCGAGA 541
QY 5711 atgccagtggtgctgctggttgaagtaaacgagagcaggtgtgctctatccagctccttact 5770
DB 542 ATGCAGGGTGGTCTGTGTGAAGAGAAAGCCAGAGGCTGTGCCTACCCAGCCCTTACTT 601
QY 5771 ggaatgaatgaggaacagacccttgcctggaagagggggaaccccttcttcttcttcctcg 5830
DB 602 GGATGAATATGGAGAAACAGACCCAGGGCTAAAGAGAGAGAAACCCACTTCAATTTATCTCG 661
QY 5831 tgagcggatcagaagctccatttgctc-tggcaacaacactgcattatagaacagattg 5889
DB 662 GGAGCGGTATCGGAAGCTGCATTTGGTCTTGGCAAAAGGACTGGATTATAGAAGAGAAATG 721
QY 5890 ctaggagccaagagactaatcagatgttattgtgattcaactgcagtcagcttgagctc 5949
DB 722 GTCGGAGCCGGGAGACTAATCAGATGCTATTTGGACTTAAGTGGAGTACTCTCTGAGATC 781
QY 5950 caactctgctcacaaga 5965
DB 782 CAGTTGTCTCAAGAGA 797

RESULT 5
BI086469 782 bp mRNA EST 20-JUN-2001
LOCUS 602849734F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:4991193 5',
DEFINITION mRNA sequence.
ACCESSION BI086469
VERSION BI086469.1 GI:14504799
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 782)
AUTHORS NIH-MGC http://imgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: ATCC
              CDNA Library Preparation: Life Technologies, Inc.
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLAM11008 row: d column: 10
              High quality sequence stop: 715.
              Location/Qualifiers
                1..782
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /clone="IMAGE:4991193"
                  /clone_lib="NIH_MGC_10"
                  /cell_line="MGC36"
                  /lab_host="DH10B"
                  /note="Organ: cervix; Vector: pCMV-SPORT6; Site:1: NotI;
                  Site:2: SalI; Cloned unidirectionally. Primer: Oligo dt.
                  Average insert size 1.5 kb. Library prepared by Life
                  Technologies."
BASE COUNT 184 a 194 c 163 g 241 t
ORIGIN

```



```

Db 85 GAGGCTGGAAAGTATTGGAGGCTTTTGTGCTCCATGTCAGGTTCACTTACATCAATA 26

Qy 6081 aaattttcttaattga 6097

Db 25 AAATATTTCTTAATGA 9

RESULT 7
BE589438
LOCUS BE589438 606 bp mRNA EST 28-AUG-2000
DEFINITION 195602 BARC 5BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION BE589438
VERSION BE589438.1 GI:9842477
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 606)
Sonstegard,T.S., Capuco,A.V., Van Tassel,C.P., Ashwell,M.S. and
Wells,K.D.
TITLE Mapping of Expressed Sequence Tags from a normalized bovine mammary
gland cDNA library
JOURNAL Unpublished (2000)
COMMENT Contact: Sonstegard TS
USDA, ARS, Beltsville Agricultural Research Center
Bldg. 200 Rm 2A, Beltsville, MD 20705, USA
Tel: 301 504 8416
Fax: 301 504 8414
Email: tads@psi.barc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCCAGTCAGCG
Plate: 120 row: N column: 11
Seq primer: ATTTAGGTGACACTATAG.
FEATURES
source
1..606
Location/Qualifiers
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="BARC 5BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled mRNA isolated from mammary
tissues at eight physiological, developmental, and disease
states."
BASE COUNT 144 a 146 c 113 g 203 t
ORIGIN
Query Match 8.1%; Score 512.2; DB 10; Length 606;
Best Local Similarity 90.4%; Pred. No. 4.2e-109;
Matches 547; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

Qy 4761 aaagcattaatgcagttgtcagtcacagagattacctgtcctcaggtctcgtatcacag 4820
Db 1 AAAGCATTAATGCAGTTTGCCATTGCACAAAGATTACCTGTTCTCAGGTCTCTGATACAG 60
Qy 4821 aaacatcgttcgtcttctatcagttgttcttctcctcaataaaatcagaagatacacca 4880
Db 61 AAACATCTGATTGCTGTTCTATCAGTTGTCTCTTAACCTGCAATCAGAAGCTACACCT 120
Qy 4881 tgccctctctatagatctgtttcatgttttgggtggtgctgtgttagcattcccatcc 4940
Db 121 TGCCCTCTATCCATAGATCTATTTCAAGTTTGGGGGGTACTGTGTAGCATTCCTCGTCC 180
Qy 4941 ttgtattgggatgacctgttgatctgcagcctcttctcagttagttcttctctataaccac 5000
Db 181 TTATACTGGGATGATGCTGTTCTTCTGCTGCGCCTTCCACAGTTAGCTTCTTCGTATAACCAC 240

```

```

Qy 5001 ctttatcttccatttgatcaccatggcacacatgcttcagatactactactacagtagac 5060
Db 241 CTTTATCTCTTCCATTGATCACCATGGCACATGCTTCAGATACCTTCCTTACCATAGAC 300
Qy 5061 acagggctaccctctgctcaggttcaagaagacagtgtaagaggctcattccgcactctct 5120
Db 301 ACAGACCTCCCTTGGCAGAGATACAGGAGAGAGTGAAGAGGCTGTTCTGTCATCTCT 360
Qy 5121 ttctttgcagaaattcttcaatatacataagtggtccctccattgggtgtagatattcctggctgg 5180
Db 361 TTCTTGGCAGAGTCTTCTCAGTATACAAAGTGGCTGATGGTGTGGTATTCTTCTGGCTGG 420
Qy 5181 tatttggtggtctcactgaagaatggtcatcaccccttatctctgctgtgctgctattgttt 5240
Db 421 TATTATGGGTCTCTTGAAGAATGGCATCATCCCGTATCTTCGCTGTGCTGCAATGTTT 480
Qy 5241 ttccactatttacttggtgtaactccgctcaggaactgcataccaattctgcagaagga 5300
Db 481 TTCCACTATTACTTGGAGTAACCTCCACCTGAAGAAGCTTTTACCAATGCTGCGAAGGA 540
Qy 5301 ggtacagtcactcttagctatctatcttacctacaaattgttctcgtctcttccag 5360
Db 541 GAATACAATGCTACTCTGTAGCTATTTATCTTACCCACAATTTGTTCTCTGCTTTCCAG 600
Qy 5361 gaata 5365
Db 601 AAATA 605

RESULT 8
BG219270/c
LOCUS BG219270 624 bp mRNA EST 21-APR-2001
DEFINITION RST39023 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION BG219270
VERSION BG219270.1 GI:13745291
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 624)
Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Dahl,T., Thornton,M., Ramachandran,R., Whittington,J.,
Lerner,L., Krashoc,D., McElligott,K., Clark,S., Mays,R., Smith,E.,
Veloaso,N., Hess,J., Cothren,K., Lo,K., Offenbacher,J., Danzig,J.
and Ducar,M.
TITLE Creation of Genome-wide Protein Expression Libraries using Random
Activation of Gene Expression
JOURNAL Nat. Biotechnol. 19 (5), 440 (2001) In press
COMMENT Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: sca@atersys.com
High quality sequence stop: 624.
FEATURES
source
1..624
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="HT1080"
/note="See 'Creation of Genome-wide Protein Expression'
Libraries using Random Activation of Gene Expression',
Nature Biotechnology, in press. Note that even though the
cell type indicated is HT1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in HT1080 under normal circumstances."
BASE COUNT 129 a 151 c 136 g 206 t 2 others
ORIGIN

```

```
Query Match 7.6%; Score 479.8; DB 11; Length 624;
Best Local Similarity 98.2%; Pred. No. 1.6e-101;
Matches 484; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 3723 attactcatgataaagaaagcagaacgaaagaaagcgtgagctgtaggtacat 3782
Db ATTACTCATGATAAAGAAAGCAGACGAAAGAAAGCTGNAGCTGNTAGTTACAT 565

QY 3783 gcgcagaagatcagctgagctgctccttacagaaaaaacttcattgaaactcataa 3842
Db CGCCAGAAAGATCATGGCTCAGATGCTGCTTACAGAAAAAATTCATTGAAACTCATAA 505

QY 3843 ctcatgtatgacatacatcagaatgcctgggaaagagattccattatgagagaag 3902
Db CTCTATGTATGCAATATACATCAGAAATGCTGGGAAGAGATTCATTATGAGGAAGAG 445

QY 3903 agcaccacagcagtcagtgactactctagaattgctttgggtcctaaacggggtccatc 3962
Db AGCACCACAGCAGTCAGTACTACTAGAAATGCTTTGGGTCTTAACGGGGTCCATCT 385

QY 3963 gttactgaaagagggtgctgacgtgcacatccttggcgaagaaagcaggggtgaaata 4022
Db GTTACTGAAAGAGGAGTGTGACGTGCATCCTTTGCCAAGAAAGACAGGAGGTGAAATA 325

QY 4023 gaaaataatgccatgtattatcgctgtgtccagaaatctactgccttaaccagcac 4082
Db GAAATAATGCCATPGGTATTATCGCTGTGTCTCCAGAAATCTACTGCCCTTAACCCAGCAC 265

QY 4083 aggggaaacccatagaaactcaggaagccctagaccacatttcagatccagac 4142
Db AGGGGAAACCCATAGAACTCTCAGGAGAACCCCTAGACCACCTTTTCATGGATCCAGAC 205

QY 4143 ttggcatatggaacttacaggaagctgtggtcatgttaatgcacgcagtgctgtggcag 4202
Db TTGGCATATGGAACCTTATACAGGAAGCTGTGTCATGTAAATGCACGCGATGTGCTGGCAG 145

QY 4203 aagtattttgaag 4215
Db AACTCCTCCGGAG 132

RESULT 9
AII92195/c
LOCUS AII92195 481 bp mRNA EST 28-OCT-1998
DEFINITION qc92e08.x1 Soares_pregnant_uterus_NBHPU Homo sapiens cDNA clone
IMAGE:1721702 3' similar to TR:015057 O15057 KIAA0349 ;, mRNA
sequence.
ACCESSION AII92195
VERSION AII92195.1 GI:3743404
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 481)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgap@remail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 563 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 351.
Location/Qualifiers
1. .481
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1721702"
/clone_lib="Soares_pregnant_uterus_NBHPU"
/sex="female"

FEATURES
source
```

```
/dev_stage="adult"
/lab_host="DH10B"
/Note="Organ: uterus; Vector: pT73-Pac; Site.1: Not I;
Site.2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5',
AACTGGAGAATTCGGCGCGCTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaldo."
```

BASE COUNT 125 a 96 c 86 g 174 t
ORIGIN

Query Match 7.6%; Score 477.8; DB 10; Length 481;
Best Local Similarity 99.6%; Pred. No. 4.6e-101;
Matches 479; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2778 cttcagattggtgcattctttaaaggatcccaataagttctgttactggttacttcagagg 2837
Db CTTCAGATTGGTGCATCTTTAATGGATCCAAATAAGTCTCTGTGTACTGGTACTTCAGAGG 422

QY 2838 tatgaactgcccagggccttttaacaagaccatctctacaaaaagccaggatttattataa 2897
Db TATGAACCTTGGCGAGGCTTTTAAACAAGACCATATCTACAAAAGCCAGGATTTGATTAAA 362

QY 2898 caataataacactaataagaagaatgcttcagctcctcatctatattgtgggtgagcgt 2957
Db CAATAATAACACTAATAAGAAGAAATGCTTCAGGTCTCTCATATATTGTGGGTGAGCGT 302

QY 2958 tatgtacctggagggtggaatgtgaccaagaagagtgcaaatgagagaatcatcac 3017
Db TATGTACCTGGAGTGGGAAATGTGACCAAGAAAGAGGTCAACATGAGAGAAATCATTCAC 242

QY 3018 ttgcttgcattgaacccatgcccacagtgccattgcccataaatttaccctgagaatgaa 3077
Db TTGCTTTGCATTGAACCCATGCCACACAGTGCCATTGCCAAAAATTTACCTGAGAATGAA 182

QY 3078 aataatgaactggcttagaagaatgtcataaacaagaagtgccacatttaagaaacccaggt 3137
Db AATAATGAACCTGGCTTAGAAGAATGTCAATAAAGTGCCACATTTAAGAAACCCAGGT 122

QY 3138 gtatcaggccatggagctttatgaactaaagatgaatcactgaagacttcaatgtac 3197
Db GTATCAGGCCATGGAGCTTTATGAACCTAAAGATGAATCACTGAAGACTTCAATATGTAC 62

QY 3198 tttatcattactccaaacccagcatagcaagctggaacatgcaagaagaaagaga 3257
Db TTTTATCATTTACTCCAAAACCCAGCATAGCAAGGCTGAACATATATGCAGAAAAGGAGA 2

QY 3258 a 3258
Db 1 A 1

RESULT 10
AW971391
LOCUS AW971391 583 bp mRNA EST 01-JUN-2000
DEFINITION EST383480 MAGE resequences, MAGL Homo sapiens cDNA, mRNA sequence.
ACCESSION AW971391
VERSION AW971391.1 GI:8161236
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 583)
Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
Quackenbush,J.
Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray

JOURNAL COMMENT

Unpublished (2000)
Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@tigr.org
Plate: 289
Seq primer: Forward.
Location/Qualifiers
1. 583
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE resequences, MAGL"
/note="Vector: pBluescriptSKm"
171 a 126 c 137 g 149 t

FEATURES source

Query Match 7.5%; Score 473.4; DB 10; Length 583;
Best Local Similarity 99.8%; Pred. No. 5.1e-100;
Matches 474; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 5678 accgaggtctcatttctctaaataacagagaatgccaggtggtcctgttgaaagtaa 5737
Db 105 AGCCGGAGTCTCATTTCTCTAAAATCAGAAATGCCAGGTGGTCTGTGTTGAAGGTAA 164
Qy 5738 accgagaggtcgtcattccagctccttacttgatgaatgagaaacagacctg 5797
Db 165 ACCCAGAGGCTGTGCTATCCAGCTCCTTACTTGGATGAATATGGAACAACAGCCCTGG 224
Qy 5798 cctgaagaggggcaaccccttcatttatctgtgagcgggtatcggaagctccatttgg 5857
Db 225 CCTGAAGAGGGCAACCCCTTCATTATCTCGTAGCGGTATCGGAAGCTCCATTGGT 284
Qy 5858 ctggaacaaactgcatttatagaagagattgtagagccaaagagactaatcagatgtt 5917
Db 285 CTGGCAACAACACTGCATTATATAGAGAGATTGCTAGGAGCCCAAGAGACTAATCAGATGTT 344
Qy 5918 atttgattcaactgacacttactgtgagctcctcaactctgctcagacaatcacaaatg 5977
Db 345 ATTGGATTCAACTGCAGTCTACTGTGAGCTCCAACTCTGCCCTCAAGACAATCACAAATG 404
Qy 5978 acqacagtagtaagaggtgatcaaaattatgaaactttctgagggctggaaagtat 6037
Db 405 ACGACAGTAGTAAGGCTGATTCAAATTTATGAAACTTTCTGAGGGCTGGGAAAGTAT 464
Qy 6038 tggagggcttttgcctccatgtccaggttcacttacatcaataataattcttaattgga 6097
Db 465 TGGAGGGTCTTTTGTCTCCATGTCAGGTTTCACTTACATCAATAAATAATTTCTTAATGGA 524
Qy 6098 gtattgctttcaattagcaaacatattgcttcacagaaagagacatatgcaa 6152
Db 525 GTATTGCTTTCAATTAGCAACATATGCTTCACAGGAAAAAGGACATAGATCCA 579

RESULT 11

AA401319 LOCUS AA401319 478 bp mRNA EST 09-NOV-1997
DEFINITION zu63d04.r1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:742663
5', mRNA sequence.
ACCESSION AA401319
VERSION AA401319.1 GI:2053683
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 478)
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin,
J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,
White, Y., Wylie, T., Waterston, R. and Wilson, R.

JOURNAL COMMENT

WashU-NCI human EST Project
Unpublished (1997)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1003 Std Error: 0.00
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 458.
Location/Qualifiers
1. 478
/organism="Homo sapiens"
/db_xref="GDB:5929590"
/db_xref="taxon:9606"
/clone="IMAGE:742663"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech Laboratories
, Inc., and primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGGCCCAATTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7m3 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bernaldo. "

FEATURES source

BASE COUNT 114 a 109 c 114 g 141 t
ORIGIN
Query Match 7.5%; Score 471.6; DB 10; Length 478;
Best Local Similarity 99.2%; Pred. No. 1.3e-99;
Matches 474; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 5173 ctggctggtatttgggtctcactgaagaatggcatcaccccttatcttcgctgtgctg 5232
Db 1 CTGGCTGTGATTGTGGGTCTCACTGAAGAATGGCATCACCCCTTATCTTCTGCTGTG 60
Qy 5233 cattgttttccactatttacttgggttaactccgctcgaggaactgcataccaattctg 5292
Db 61 CATGTGTTTCCACTATTACTTGGGTAACTCCGCTGAGGAAGTGCATACCAATTCTG 120
Qy 5293 cagaagagagtagcagtcgactctgtagctatctatttaccatacaatttgcctgc 5352
Db 121 CAGAAGGAGAGTACAGTGCACCTCTGTAGCTATCTATCTTTACCTGCAAAATTTGTTCCCTGC 180
Qy 5353 tcttccaggaatattgggatactgaagcccttgctccagagtggtggtcagatcctg 5412
Db 181 TCTTCCAGGAATATTTGGGATACTGTAGGCCCTTGTCTCCAGAGGTGGTGTGAGATCTGT 240
Qy 5413 ccttactaaactgttttgaagcaaaaaaacacgtggtcaggtacccctagagaaagaaata 5472
Db 241 CCTTACTAACTGTTTGAAGCAAAAACACCGTGGTCAGGTACCTAGAGAAAGAAATA 300
Qy 5473 gttttagagcttcctgatgactatagctgcctcctgaaatcaagcttctcattcaggt 5532
Db 301 GTTTGATAGAGCTTCTGTGATGACTATAGCTGCCCTCTCTCTCTCTCTCTCTCTCTCTCT 360
Qy 5533 gccacaggtctgcagatgatgagcgaagatcctgctcctgtcctctgttctgtgggcta 5592
Db 361 GCCCAGCGTCTGCAGATGATGAGCGAAACAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 420
Qy 5593 tactatattctcagaacatttgcctcaggaattgtgaacgggaagaggttgagc 5650
Db 421 TACTATTGTTCTCAGAACATTTGCTGCCAGGAATTTGTGAACGGGGAAGAGGTGGAGC 478

RESULT 12

AW291190/c
 LOCUS
 DEFINITION
 UI-H-BI2-agb-g-08-0-UI.s1 NCI_CGAP_Sub4 Homo sapiens cDNA clone
 IMAGE:2723895.3', mRNA sequence.

ACCESSION
 VERSION
 KEYWORDS
 SOURCE

AW291190.1 GI:6697826
 EST.
 human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 522)

REFERENCE

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

AUTHORS

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

TITLE

Tumor Gene Index

JOURNAL

Unpublished (1997)

COMMENT

Contact: Robert Strausberg, Ph.D.
 Email: cgapb-r@mail.nih.gov

The sequence contained an oligo-dT track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. cDNA library Preparation: M.B. Soares Lab Clone Distribution:
 NCI-CGAP clone distribution information can be found through the
 I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Seq primer: M13 Forward

POLYA=Yes.

FEATURES

Location/Qualifiers

1..522
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2723895"
 /clone_lib="NCI_CGAP_Sub4"
 /lab_host="DH10B (Life Technologies)"
 /note="vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; The
 NCI_CGAP_Sub4 library is a subtracted library derived from
 the NCI_CGAP_Sub2 library which is a subtracted library
 derived from the NCI_CGAP_Sub1 library, which is a
 subtracted library derived from BI. BI constitutes a
 mixture of 21 normalized or subtracted NCI_CGAP
 libraries: NCI_CGAP_Co4, NCI_CGAP_Co16, NCI_CGAP_Pr28,
 NCI_CGAP_Co10, NCI_CGAP_Co16, NCI_CGAP_Kid5,
 NCI_CGAP_Kid12, NCI_CGAP_Kid3, NCI_CGAP_Kid11,
 NCI_CGAP_Lym2, NCI_CGAP_Br2, NCI_CGAP_Co8, NCI_CGAP_CLL1,
 NCI_CGAP_Le12, NCI_CGAP_Brn23, NCI_CGAP_Lu5,
 NCI_CGAP_Lu24, NCI_CGAP_Brn23, NCI_CGAP_Lu5,
 NCI_CGAP_Lu24, NCI_CGAP_Lu19, NCI_CGAP_GC4, NCI_CGAP_GC6,
 NCI_CGAP_Brn25. These 21 libraries were pooled and a
 single-stranded DNA preparation of the resulting mixture
 was used as a tracer in a subtractive hybridization with
 a driver whose composition is detailed below:
 NCI_CGAP_Kid3 pool 1 : LLAM 3334-3337, 3682-3683,
 3798-3803 (IMAGE Clonides 1322376-1323911,
 1456008-1456775, 1500552-1502855) NCI_CGAP_Kid5 pool 1 :
 LLAM 3338-3342, 3722-3725, 3776-3778 (IMAGE Clonides
 1323912-1325831, 1471368-1472903, 1492104-1493255)
 NCI_CGAP_Lu5 pool 1 : LLAM 3575-3582, 3851-3854 (IMAGE
 Clonides 1414920-1417991, 1520904-1522439) NCI_CGAP_GC4
 pool 1 : LLAM 3164-3167, 3716-3720, 3733-3735 (IMAGE
 Clonides 1257096-1258631, 1469064-1470983, 1475592-1476743
) NCI_CGAP_Pr22 pool 1 : LLAM 2457-2459, 2758-2759
 3062-3068 (IMAGE Clonides 985608-986759, 1101182-1101959,
 1217928-1220615) NCI_CGAP_Co10 pool 1 : LLAM 2644-2653,
 2871-2872 (IMAGE Clonides 1057416-1061255, 1144584-1145351
) Subtraction was performed as previously described
 [Bonaldi, Lennon & Soares (1996): Normalization and
 Subtraction: Two Approaches To Facilitate Gene Discovery.
 Genome Research 6, 791-806.]
 TAG_LIB=NCI_CGAP_Co4
 TAG_TISSUE=colon
 TAG_SEQ=CTTCG"

BASE COUNT

137 a 125 c 100 g 160 t

ORIGIN

Query Match 7.2%; Score 454; DB 10; Length 522;

Best Local Similarity 98.0%; Pred. No. 1.7e-95;

Matches 492; Conservative 0; Mismatches 5; Indels 5; Gaps 3;

QY 5596 tatgtttcagaacatttgccaggaaattgtgaacgggggagaggttgagcttga 5655

|||||

Db 513 TATGTCTCTCAGAACATTTGCTGCCAGGAAA-TGTGAACGGGAGAGTTGGAG---TTCA 458

|||||

QY 5656 tttttcagcaacttctggtggagccgaggtctgcattttcctaaaaatcagagaatgcc 5715

|||||

Db 457 TTTTTCACGCACACTTCACCTGTGGAGCCGAGTCTGCATTTTCTTAAAAATCAGAGAATGCC 398

|||||

QY 5716 gactggtccctggtgaaggttaaagccagagctgctcctatccagctccttacttgatg 5775

|||||

Db 397 GAGTGTCTCTGTTGAAGGTAAAGCCAGAGGCTGTCTATCCAGCTCCTTACTTTGGATG 338

|||||

QY 5776 aatatggagaacacagaccctggcctgaaggggcaaccccttcatttctcgtgagc 5835

|||||

Db 337 AATATGGAGAAACAGACCCCTGGCTGAAGAGGGGCAACCCCTTCATTATCTCGTGAGC 278

|||||

QY 5836 ggtatcggagctccatttggcttgctggcaacacactgcattatagagagattgctagga 5895

|||||

Db 277 GGTATCGAAGCTCCATTTTGGTCTGCAACAACACTGCATTATAGAGAGATTGCTAGGA 218

|||||

QY 5896 gccacagactaatcagatgtatttggtattcactgagctgagcttactgagctccaactc 5955

|||||

Db 217 GCCAAGAGACTAATCAGATGTTATTGGATTCAACTGGCAGTTACTGTGAGCTCCAACCTC 158

|||||

QY 5956 tgctcaagacaatcacaaatgacacagtagtaaaaggtgattcaaaatattggaataac 6015

|||||

Db 157 TGCCTCAAGACATCAACAATGACGACAGTAGTAAGAGCTGATTCAAATTTATGGAAAAAC 98

|||||

QY 6016 ttctgagggctgggaaagtattggagggtcttttctcctcattgctccaggttccacttaccat 6075

|||||

Db 97 TTTCTCAGGGCTGGAAAGATTGGA-GGTCTTTTGTCTCCATGTCCAGGTTCACTTACAT 39

|||||

QY 6076 caataaaattttcttaatgga 6097

|||||

Db 38 CAATAAAATATTCTTAAATGGA 17

|||||

RESULT 13

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Sus scrofa

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

1 (bases 1 to 522)

AUTHORS

Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,

Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.

and Keese,J.W.

Design and use of two pooled tissue normalized cDNA libraries for

EST discovery in swine

Unpublished (2000)

Contact: Smith TPL

USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366

Fax: 402 762 4390

Email: smathe@mail.marc.usda.gov

Single pass sequencing. Bases called and alt.trimmed with phred

v0.980904.e. Vector identified by cross_match with the -minscore 18

and -minmatch 12 options.

PCR primers

FORWARD: AGGAAACAGCTATGACCAT

Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine
Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390

FEATURES SOURCE

BASE COUNT

ORIGIN

•

Query Max
Best Loc

Matches

Qy 4919

Db 1

Qy 4979

Db 61

0v 5039

1	10
2	20
3	30
4	40
5	50
6	60
7	70
8	80
9	90
10	100

121 of

QY 5099

Db 181

Qy 5159

Db 241

Ov 5219

103 201

TOC 22

QY 5279

Db 361

Qy 5339

Db 421

Qy 5399

Dh 481

4
2
2

RESULT 1

```

BE650873/c
LOCUS      BE650873      486 bp      mRNA      EST      06-SEP-2000
DEFINITION UI-M-BH3-ata-a-07-0-UI.r1 NIH_BMAP_M_S4 Mus musculus cDNA clone
ACCESSION  BE650873
VERSION    BE650873.1 GI:9976697
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 486)
AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL   Genome Res. 6 (9), 791-806 (1996)
MEDLINE   9704477
COMMENT   Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEst@mail.nih.gov
            CDNA Library Preparation: M.B. Soares Lab Clone distribution:
            Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It
            should be noted that Bento Soares is generating a small number of
            additional specialized non-redundant arrays of BMAP cDNAs whose
            availability will be considered under appropriate and limited
            collaborative arrangements
Seq primer: M13 Reverse.
            Location/Qualifiers
                1..486
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="UI-M-BH3-ata-a-07-0-UI"
                /clone_lib="NIH_BMAP_M_S4"
                /dev_stage="27-32 days"
                /lab_host="DH10B (Life Technologies)"
                /note="vector: pMT30-Pac (Pharmacia) with a modified
                polylinker; Site_1: Not I; Site_2: Eco RI; The
                NIH_BMAP_M_S4 library is a subtracted library of a series,
                ultimately derived from a mixture of individually tagged
                normalized libraries from ten regions of the mouse brain
                (cerebellum, brain stems, olfactory bulbs, hypothalamus,
                cortex, amygdala, basal ganglia, pineal gland, striatum,
                hippocampus) after a series of subtractions to reduce the
                representation of cDNAs from which ESTs had already been
                generated. The following serially subtracted libraries
                were generated in this process: NIH_BMAP_M_S4,
                NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S3.1,
                NIH_BMAP_M_S2, NIH_BMAP_M_S1. The subtracted library
                (NIH_BMAP_M_S4) was constructed as follows: PCR amplified
                cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and
                NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived
                was used as a driver in a hybridization with a pool of
                the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1
                libraries in the form of single-stranded circles. The
                remaining single-stranded circles (subtracted library)
                was purified by hydroxyapatite column chromatography,
                converted to double-stranded circles and electroporated
                into DH10B bacteria (Life Technologies) to generate the
                NIH_BMAP_M_S4 library. This procedure has been previously
                described (Bonaldo, Lennon and Soares, Genome Research
                6:791-806, 1996)"
BASE COUNT      129 a      117 c      95 g      145 t
ORIGIN

```

FEATURES

source

Search completed: March 1, 2002, 14:46:36
Job time: 8041 sec

Query Match 6.5%; Score 410.8; DB 10; Length 486;
Best Local Similarity 90.3%; Pred. No. 2.3e-85;
Matches 439; Conservative 0; Mismatches 47; Indels 0; Gaps 0;